cGraw-Hill Encyclopedia

McGRAW-HILL BOOK COMPANY INC.

NEW YORK CHICAGO SAN FRANCISCO DALLAS TORONTO LONDON

of Science and Technology

AN INTERNATIONAL REFERENCE WORK

IN FIFTEEN VOLUMES INCLUDING AN INDEX

VOLUME 4 DAC ENS



(LEFT) A single crystal of antimony, cleared and deeply etched by an acid mixture along (111) planes (*Bell Telephone Laboratories, Inc.*) (RIGHT) Sahara desert sand dunes in southern Libya (*Aero Service Corp.*)

Library of Congress Catalog Card Number 60 11000

Suggestions to the Readers

The basic plan of the Encyclopedia is explained here in order to facilitate its use

The subject matter of the various disciplines or branches of science and technology is organized systematically a general article provides a broad survey of the field and a number of separate articles alphabetically arranged cover its main subdivisions and more specific aspects

Cross references guide the reader from the general articles to the other articles into which the subject is subdivided and from these to articles on more highly specialized phases of the subject The cross references—there are about 50 000 of them—are printed in small capital letters so that they can be easily recognized By means of the cross references a reader may find his way from ELECTRICAL ENGINEERING through ELECTRONICS and VACUUM TUBE to ELECTRON MOTION IN VACUUM OF ELECTRON EMISSION Or following another line of cross references the reader would be led to ELECTRIC POWER SYSTEMS TRANSMISSION LINES ELECTROMAGNETIC WAVE and so on

In general each article begins with a definition of the title that states its scope and coverage Usually only the scientific or technological sense is discussed Most of the articles after this statement go on to increasingly complex and detailed considerations A reader thus needs to proceed only as far as his inclinations and requirements dictate

The Index Volume 15 should be consulted to locate the discussion of topics covered in the Encyclopedia but not given in separate entries

Every phylum class and order in the plant and animal kingdoms is allotted a separate article Many of the more common families genera and species are covered either in one of the order articles or in a separate article under its own scientific or common name

The adjectives electric and electrical are used in the following senses Electric—containing producing arising from actuated by or carrying electricity or capable of doing so as for instance electric generator electric motor electric wiring Electrical—related to pertaining to or associated with electricity but not having its properties or characteristics as for example electrical code electrical engineering

Words used as titles are wherever possible given in the singular to permit a consistent alphabetic arrangement Titles are alphabetized by word and not by letter for example

Earth sciences
Earth tides
Earthmover
Earthquake

A word used as a noun precedes the same word used adjectivally thus

Mercury (element)
Mercury (planet)
Mercury battery

or

Circuit, electronic
Circuit breaker

Hyphened terms are alphabetized as single words for example

Animal virus
Animal feed composition

Most of the longer articles contain bibliographies citing useful sources of further information For additional bibliographical citations the reader should refer to related articles (as indicated by the cross references in the article) Bibliographies are placed at the ends of articles or sometimes at the ends of major sections in long articles

A list of initials and names of the contributors to the Encyclopedia is to be found in Volume 15 This list will permit quick identification of a contributor s initials after an article Immediately following this list is a second list of encyclopedia contributors with their affiliations and the titles of articles each has written for the Encyclopedia

McGraw-Hill Encyclopedia of Science and Technology

D *Dace to Dysprosium*

Dace

Any of several minnows of the family Cyprinidae The name has no taxonomic standing but usually is applied to small, dark, fine scaled minnows living in cold running water, although some species



The horned dace, *Semotilus atromaculatus*, length to 10 in (*From E L Palmer, Fieldbook of Natural History, McGraw-Hill, 1949*)

live in lakes They feed upon plankton and insects and are important forage for such game fish as trout and smallmouth bass They are commonly used as live bait for these and other fishes *See* ACTINOPTERYGII [J D B]

Dacite

Aphanitic (very finely crystalline or glassy) rock of volcanic origin, composed chiefly of sodic plagioclase (oligoclase or andesine) and free silica (quartz or tridymite) with subordinate dark colored (mafic) minerals (biotite amphibole or pyroxene) If alkali feldspar exceeds 5% of the total feldspar, the rock is a quartz latite As quartz decreases in abundance dacite passes into andesite Thus dacite is roughly intermediate between andesite and quartz latite. *See* ANDESITE [C A C]

Dairy machinery

Milking machines, cream separators coolers, pasteurizers, homogenizers churns and other items of equipment used in the extracting and processing of milk and milk products A major consideration in this equipment is ease of cleaning and prevention of contamination by dirt oil, soluble metals and other foreign material Stainless steel is highly satisfactory for direct contact with milk because if properly used, it does not affect flavor and is corrosion resistant, even to highly corrosive cleaning solutions (*see* STAINLESS STEEL) Surfaces in contact with the milk should be of one-piece drawn construction, where possible, to eliminate sharp corners and joints which can harbor milk residue and bacteria Industry standards have been adopted for design construction, and operation of sanitary equipment

The science of hydraulics is basic to the fluid flow processes of pumping piping agitating, centrifuging and high pressure homogenizing (*see* CENTRIFUGATION, HYDRAULICS) Electric motors are used to supply power for these processes Heating is by electricity for small equipment or by boilers furnishing hot water or steam for large equipment Refrigeration (for storage and for ice cream production) and heating (for pasteurization and production of evaporated and dried milk) require the major portion of the energy and involve the application of thermodynamic principles *See* REFRIGERATION THERMODYNAMIC PRINCIPLES

Milking machines. These extract milk from the cow's udder and deliver it into a small adjacent container or, in the case of pipeline milking systems, directly into a central cooling tank (Fig 1)

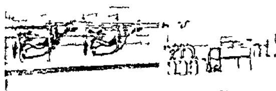


Fig 1 *Milking machines, milking parlor, and milk house (DeLaval Separator Company)*

Teat cups fit over the cow's teats and extract milk by intermittent vacuum action on a flexible inner wall called an inflation The action is more efficient than that of manual milking With proper techniques most cows can be milked in 5 minutes hand milking may require several times as long

A pump developing a vacuum of 10–15 in Hg usually driven by an electric motor and a vacuum storage tank are required to actuate the milker (*see* VACUUM PUMP) Intermittent action of 45–48 cycles per minute is secured by an automatic pulsator control valve Transfer of milk to a central point by pipeline is also accomplished by vacuum

Thorough cleansing of all milk contacting parts is essential to the production of milk of an acceptably low bacteria count (*see* MILK) Washing should be done immediately after using by first drawing through cold water, then hot water suds (using a wetting agent), rinsing with boiling water, and dismantling (*see* SURFACE ACTIVE AGENT) A sterilizing solution may be run through the reassembled machine just before milking To meet public health ordinance requirements milking m

19 control panel with thermal limit recorder controller, hot water controller, start and stop buttons
flexible control cable
11 pasteurized milk to regenerator
9 cable to controller bulb
9A indicator thermometer
electric wiring (controller to solenoid valve)
air line
air supply
15 cold pasteurized milk out
12 4
10 flow diversion valve
flexible control cab
air line
2 float tank
1 raw milk in
14 coolant in
3 raw milk to regenerator from float tank
diverted milk flow line into float tank
17 water heating and circulating unit
8 holder tube
6 milk to final heater
steam supply
18 hi flow steam control valve
5 milk timing pump
raw milk
pasteurized milk
hot water
coolant

Fig 2 Milk pasteurizer Flow chart for HTST (high-temperature short time) plate pasteurizer Numbers (1–19) show sequence of flow through the equipment (Creamery Package Manufacturing Company)

chinery must be cleaned and stored in a sanitary milk house of approved specifications

Milk pasteurizer This type of dairy equipment is used to heat milk to a predetermined temperature and hold it there long enough to kill the objectionable organisms which may be present Satisfactory pasteurization may be obtained at time-temperature combinations ranging from 30 min at 143°F to 15 sec at 160°F Flash pasteurization at 230°F followed by discharge into a vacuum chamber is also used in the vacuum process Batch type pasteurizers are heated externally with hot water and require thorough agitation to heat the milk uniformly Continuous-flow pasteurizers use a high temperature short time (HTST) cycle such as 0 25 min at 160°F and can use regeneration (heat transfer from the hot outgoing milk to the cool incoming milk) to improve efficiency (Fig 2)

Cream separators. Mechanical centrifuges used to extract butterfat from milk are called cream separators Milk normally contains from 3 to 5% butterfat in the form of globules which are lighter in density than the basic fluid (skim milk) These particles rise to the surface when fresh milk is al

[illegible]

airtight bowl turning from 6 000 to 10 000 rpm and

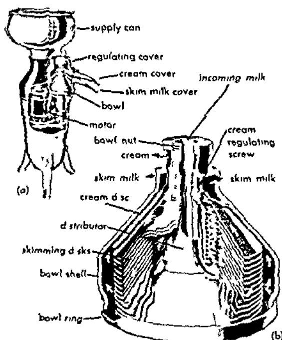


Fig 3 (a) Cream separator, electric motor driven (b) Cross section of separating bowl (DeLaval Separator Company)

subjecting the milk to radial forces up to 500g The bowl contains stacked disks in the form of inverted cones which divide the milk into thin layers and provide the fluid friction necessary to bring the milk up to the needed rotational speed The angle of the inverted cone is such that the lighter cream particles tend to be forced up to the center where they are drawn off and the heavier skim milk flows down toward the outside of the bowl and then to the top where it is collected An efficient separator will leave only 0 10% butterfat in the milk The bowl should be dismantled for thorough washing after using since sanitation requirements are stringent for all items of dairy equipment For machinery and equipment used in other phases of the dairy industry *see* BUTTER CHEESE ICE CREAM *see also* CATTLE PRODUCTION DAIRY [C B R]

Bibliography A W Farrall *Dairy Engineering* 2d ed 1953

D'Alembert's paradox

A theorem in fluid mechanics which states that no forces act on a body moving at constant velocity in a straight line through a large mass of incompressible inviscid fluid which was initially at rest or in uniform motion) This seemingly paradoxical theorem can be understood by first realizing that inviscid fluids do not exist If such fluids did exist there would be no internal physical mechanism for dissipating energy into heat hence there would be no force acting on the body because work would then be done on the fluid with no net increase of energy in the fluid

The viscosity of many fluids is very small but it is essential in explaining the forces that act on bodies moving in them The action of viscosity creates a rotation of the fluid particles that come near the surface of a moving body This vorticity as it is called is convected downstream from the body so that the assumption of irrotationality of the fluid motion made in the proof of D Alembert s theorem does not correspond to reality for any known fluid For a winglike body this viscous action sets up a circulation around the body which creates a lifting force Viscosity also gives rise to tangential stresses at the body surface called skin friction which result in a drag force on the body The work done on the fluid by moving a body through it shows up first as kinetic energy in the wake behind the body which is gradually dissipated into heat by further action of viscosity

D Alembert s theorem does not preclude the possibility of a couple acting on the body and in fact the irrotational inviscid fluid theory does predict such a couple This couple is almost always such as to cause the body to present its greatest projected area in the direction of motion *See* FLUID-FLOW PRINCIPLES [A E BR]

D'Alembert's principle

This principle states that the resultant of the external forces F and the kinetic reaction acting on a body equals zero The kinetic reaction is defined as the negative of the product of the mass m and the acceleration a The principle is stated by the equation

$$F - ma = 0$$

While D Alembert s principle is merely another way of writing Newton s second law it has the advantage of changing a problem in kinetics into a problem in statics The techniques used in solving statics problems are then applicable and may provide relatively simple solutions to some problems in dynamics D Alembert s principle is especially useful in problems involving constraints (*see* CONSTRAINT)

If D Alembert s principle is applied to the plane motion of a rigid body the techniques of plane statics can be used The principal advantage of this approach is that in a dynamics problem the torques must be calculated about a fixed point or about the center of mass while in statics torques can be calculated about any point [P W S]

Bibliography J W Campbell *An Introduction to Mechanics* 1947 R J Stephenson *Mechanics and Properties of Matter* 1952

Dallis grass

A general term for a genus of grasses of which the most important species is the deeply rooted perennial *Paspalum dilatatum* Dallis grass is widely used in the southern United States mostly for pasture and remains productive indefinitely if well managed Dallis grass does best on fertile soils and responds to lime and fertilizer On heavier soils it remains green throughout the winter unless checked by heavy frosts Seed production is hampered by infection with ergot a fungus that invades developing seeds and produces purplish black horny bodies Ergot bearing seed heads are very toxic to livestock whether in pasture or in hay and Dallis grass must be so managed as to prevent consumption of infected heads by livestock *See* GRASS CROPS [H B S]

Dalton's law

A law of physics stating that the total pressure exerted by a mixture of ideal gases under equilibrium equals the sum of the partial pressures of the constituent gases as observed by John Dalton (1766–1844) Dalton s law also called the law of additive pressures can be stated mathematically as

$$p_m = p_a + p_b + p +$$

where p_m is the pressure produced by the mixture and p_a p_b p are the partial pressures of the several gases in the mixture Partial pressure is the pressure exerted by a constituent gas that occupies the whole volume occupied by the mixture at the same temperature and pressure in the absence of the other gases *See* AVOGADRO S LAW GAS THERMODYNAMIC PRINCIPLES [G A H]

Dam

A structure which acts as a barrier flow of water in an open channel may be classified by

sion dams divert water from a stream navigation dams raise the level of a stream to increase the depth for navigation purposes power dams raise the level of a stream to create or concentrate head for power purposes and storage dams store water for municipal and industrial use irrigation flood control river regulation recreation or power production A dam serving two or more purposes is called a multiple purpose dam

Another classification commonly used is based on the material used in the construction such as masonry concrete earth rock timber and steel Most dams now are built of either concrete or earth materials

Concrete dams Concrete dams may be either gravity or arch type Gravity dams as the name implies depend on weight for stability against over turning and for resistance to sliding on their foundations An arch dam (Figs 1 and 2) acts as a horizontal arch and most of the horizontal thrust from the water pressure against the upstream face of the dam is transmitted to the abutments of the dam The hollow type of concrete dam includes the slab and buttress or Ambursen type the round or diamond head buttress type the multiple-arch type (Fig 3) and the multiple dome type These hollow dams are all gravity types as they depend upon the weight of the structure plus the vertical component of the water pressure against the sloping upstream face of the dam to resist overturning and slipping on the foundations

Forces acting on concrete dams The principal forces acting on a concrete dam are (1) vertical forces resulting from the weight of the structure the vertical component of the water pressure against the upstream and downstream faces of the dam and the vertical component of earthquake accelerations (2) horizontal forces resulting from the horizontal component of the water pressure against the upstream and downstream faces of the dam and the horizontal component of earthquake accelerations (3) temperature stresses (4) pressures from silt and earth fills against the structure (5) ice pressures and (6) uplift pressures under the base of the structure



Fig 1 Aerial view of Hungry Horse Dam a concrete arch structure on a tributary of the Columbia River in Montana (U S Bureau of Reclamation)

The forces caused by earthquakes are usually assumed to be equal to 0 10 or 0 15 times the force of gravity acting in horizontal and vertical planes In addition to the forces produced directly by earthquakes on the mass of the dam forces are produced by the action of the earth movement on the water in the reservoir and the tailrace The water adjacent to the dam resists the movement of the dam and increases the forces acting on it The problem of calculating the water pressure against a dam resulting from earthquakes has been solved by H M Westergaard Hydrodynamic theory was used to attain the solution

Stresses resulting from temperature changes are usually disregarded in the design of concrete gravity dams but must be taken into account in analyzing arch dams

Pressures caused by silt deposited in the reservoir adjacent to the dam are assumed to depend upon the rate of deposition If the silt is deposited at a rapid rate the silt pressures are assumed to act as a fluid pressure and the weight of the silt and water is assumed to be from 100 to 120 lb/ft^3 If the silt is deposited at a slow rate over an extended period of years the silt is assumed to settle and consolidate and the silt load to act in a vertical direction only

The pressure exerted against a dam by ice is generally considered to result from the thermal expansion of the ice sheet The pressures may vary with the rate and magnitude of the temperature rise and with the thickness of the ice sheet A minimum thrust of 10 000 lb per lineal foot for ice pressure against a dam is frequently considered by designers

The uplift pressure under the base of a concrete gravity dam is generally determined by assuming that the dam is impervious and that water from the reservoir passes through the pervious foundation material at a uniform rate of flow The effects of grout curtains or other cutoff walls and of drains under the dam are taken into account in determining the uplift pressures under the dam The uplift pressures may be found by means of an electric analogy or a membrane analogy

Allowable concrete stresses Stresses do not ordinarily control the design of concrete dams of the gravity type In arch dams however the stresses do control the design of the structure The stresses adopted for the design of dams are usually conservative because these structures must have a high degree of safety as well as a long useful life The

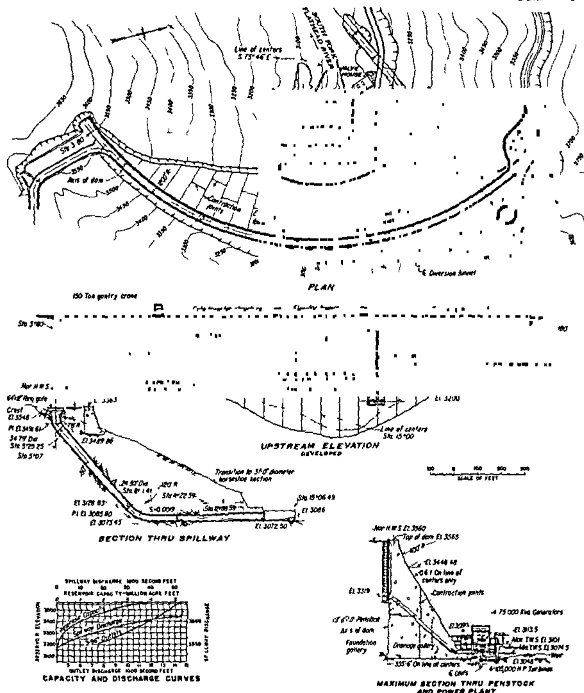


Fig 2 Plan and sections of Hungry Horse Dam (U S Bureau of Reclamation)

U S Bureau of Reclamation uses an allowable working stress of 1000 lb/in² in compression for mass concrete in major dams In calculating the stability of a gravity dam it is customary to take into account the shear strengths of the concrete and the foundation rock which are determined by actual tests of the materials

Stresses in arch dams (Fig 2) are usually analyzed by the trial load method This method is complex and takes considerable time and effort The method takes into account the stresses caused by twisting and beam action The dam is considered to be made up of systems of arches and cantilever beams The water load is divided between the structural systems by trial so that the deflections at conjugate points in the systems are equal

Concrete temperature control Temperature control of mass concrete is one of the most important problems in the design and construction o
dams The changes in volume resulting fro
trolled temperatures would produce ten
that would cause cracks and might a

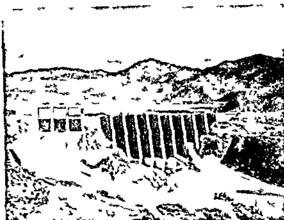


Fig 3 Bartlett Dam a multiple arch structure on the Verde River in Arizona (U S *Bureau of Reclamation*)

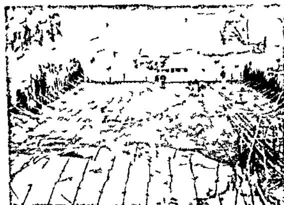


Fig 4 Cooling water pipes laid at 2½-ft centers on a prepared rock foundation surface and adjacent completed concrete l ft (U S *Bureau of Reclamation*)

bility and safety of the structure To prevent this the temperature is controlled by using special cements having low heat of hydration by using less cement in the concrete by precooling the aggregates and by circulating cold water through pipes embedded in the concrete (Fig 4)

Contraction joints are formed in the mass concrete of the dam to relieve the tensile stresses resulting from contraction of the concrete and thus prevent cracking After the concrete in the dam has cooled and the maximum shrinkage has taken place the contraction joints (Fig 5) are filled with cement grout to join the entire structure into one continuous mass The grouting mixture of water and portland cement is forced into the contraction joints by means of an embedded pipe system having outlets spaced uniformly over the faces of the contraction joints

Earth dams Earth dams have been used for the storage of water since the early civilizations Improvements in earth materials techniques and particularly the development of modern earth handling equipment have brought about a wider use of this type of dam and today as in primitive times the earth embankment is the most common type of dam (Figs 6 and 7) Earth dams may be built of loose rock, gravel sand silt or clay in various combinations

Earth dams may be classified generally as homogeneous embankments zoned embankments or rock fill embankments A homogeneous embankment consists of uniform material throughout A zoned embankment has an inner impervious section supported by upstream and downstream sections of pervious materials that sometimes include rock at the downstream toe A rock fill embankment consists entirely of rock except for an impervious membrane of reinforced concrete steel timber or asphalt on the upstream face or in the interior of the dam or a thin impervious interior core of compacted earth

Earth fill embankments whether homogeneous or zoned are usually constructed in compacted layers The layers are spread to the desired thickness and at the desired moisture content and then compacted by heavy sheepsfoot or other type rollers to secure the proper density Some earth fill embankments are placed by hydraulic fill methods

The impervious zone of an earth dam must satisfy the following criteria (1) percolation through the impervious zone or along the contacts



Fig 5 Block method of construction used on a typical concrete dam Note galleries and keyed contraction joint surfaces (U S *Bureau of Reclamation*)



Fig 6 Aerial view of Anderson Ranch Dam, a zoned earth fill structure on the South Fork of the Boise River in Idaho (U S *Bureau of Reclamation*)

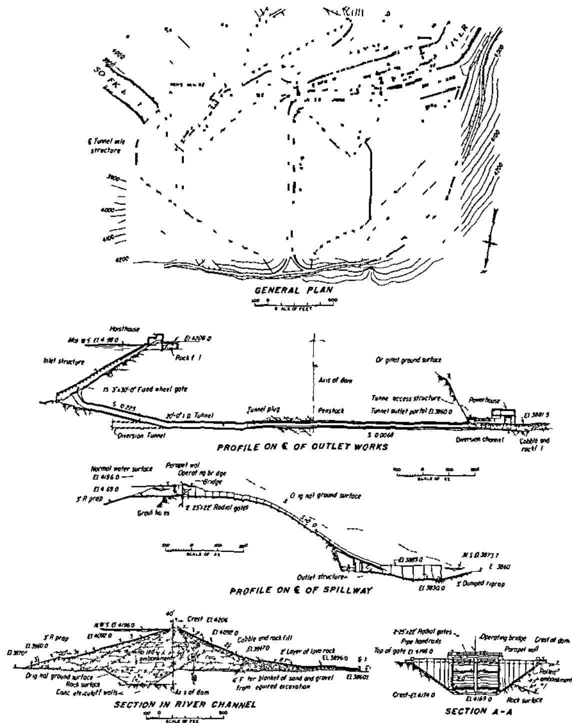


Fig 7 Plan and sections of Anderson Ranch Dam (U S Bureau of Reclamation)

with the foundations or adjoining structures must be prevented (2) settlement and consolidation caused by the weight of the superimposed material must be minimized (3) maximum practical shearing strength must be developed and (4) the material in the impervious zone must not soften excessively when it becomes saturated

These requirements can be met by proper selection of materials proper compaction and the control of moisture content During construction routine tests are made of the density and moisture content of the earth material in the embankment to ensure that the requirements are satisfied

The pervious zones of an earth dam should be free from large voids be free draining and have a high shearing strength They should not be subject to excessive settlement or consolidation under the weight of the superimposed fill These

ments can be met by selecting the proper materials and compacting them to the maximum practical density

Selection of a dam site The selection of a site for a dam depends upon many factors including foundation conditions width of river depth to bed rock width of canyon or valley topography storage capacity of reservoir and accessibility Other factors may be cost of right of way necessity for relocating railroads highways and buildings and the proximity of sources of suitable materials for concrete and for earth or rock embankments In selecting the site for a storage dam the objective is to select the site where the desired amount of storage can be most economically developed Power dams must be so located as to develop most economically the desired head and usually storage as well For a diversion dam the location of the site must be considered in conjunction with the location and elevation of the outlet canal or conduit Site selection for dams for improvement of river navigation (canalization) involves special factors such as the navigable depth and width desired slope of the river channel natural river flow amount of bank protection and channel dredging and the locations of other dams on the system

Selection of type of dam The selection of the type of dam for a particular site is made on the basis of the estimated costs of various types The most important factor is the condition of the foundation In general a hard rock foundation is suitable for any type of dam provided the rock has no unfavorable jointing and there is no danger of movement in existing faults and provided the foundation can be adequately sealed Rock foundations of high quality are essential for arch dams because the abutments have to resist the thrust of the water pressure against the face of the dam Rock foundations are necessary for all medium and high concrete dams Buttress and similar types of concrete dams because of their lighter weight may be built on foundations that cannot support a solid concrete gravity dam An earth dam may be built on almost any kind of foundation if adequate precautions are taken in its construction

Availability and cost of suitable construction materials frequently determine the most economical type of dam A concrete dam requires adequate quantities of suitable concrete aggregates while an earth dam requires adequate quantities of both pervious and impervious earth materials If impervious materials are limited and sufficient rock is available a rock fill dam is sometimes built

Diversion of river Construction of a dam necessitates unwatering the site so that the foundation may be prepared properly for the structure and so that the materials in the structure may be placed "in the dry" The stream may be diverted around the site through one or more tunnels passed through openings in the dam passed over the tops of low sections of a partially completed concrete dam or passed through or around the construction area by means of one or more flumes Sometimes the diversion of a river is conducted in two or more stages using a different method for each stage

In constructing long dams the stream may be diverted through one part of the site while the dam is being built in the other part After the first section of the dam has been completed flow may be passed through openings in the completed section and the second part of the site may then be unwatered by means of cofferdams to permit completion of the dam Usually temporary cofferdams are built both upstream and downstream from the site to keep the water out of the foundation area

Foundation treatment The foundation of a dam must provide a stable support for the structure under all conditions of operation and must prevent excessive leakage For concrete dams of either the gravity or arch type all loose and unsound material must be removed from the foundation so that the concrete structure will be on sound rock All seams fissures crevices joints or faults in the foundation rock must be filled with grout usually a mixture of portland cement and water to prevent harmful erosion excessive uplift pressures under the dam and loss of water The grout is forced into the voids in the foundation rock through holes drilled into the rock The grouting usually is performed in two stages The first stage is done before any concrete has been placed in the dam using low grouting pressures to avoid damaging the rock foundation by hydraulic jacking action This stage provides a general consolidation of the foundation rock near the surface and all major seams and crevices near the surface are filled After sufficient concrete has been placed in the dam to permit higher grouting pressures to be used deep grout holes are drilled into the foundation rock along a line extending across the channel near the upstream face of the dam and these holes are filled with grout under high pressure This provides a watertight barrier to prevent the seepage of water under the structure

In most concrete dams a line of drainage holes is drilled into the foundation rock a short distance downstream from the high pressure grout curtain These holes intercept any water that might leak past the grout curtain and thus prevent high uplift pressure under the downstream part of the dam

Concrete dams are keyed into the solid rock foundation to provide additional resistance against sliding The keys for arch dams are radial so that the thrust is at right angles to rock surfaces

For earth dams built on rock foundations the usual practice is to provide a grout curtain similar to that used for concrete dams If the foundation rock is badly fractured a trench is carefully excavated into the rock along the line of the grout curtain and is filled with concrete to form an anchor for the grout pipes through which the grouting is done to provide a barrier to seepage

Earth dams frequently are built on foundations consisting of gravel sand silt or clay or some combination thereof To prevent excessive seepage of water through these materials a trench is usually

excavated down to bedrock or to impervious material if excavation to bedrock is impractical and this trench is filled with compacted impervious materials to form a watertight cutoff If such a cutoff is impractical because of the depth to bedrock or to impervious material other means must be provided to increase the percolation distance under the dam and thus prevent the seepage water from removing fine material from the foundation

Even with a well constructed cutoff there may be some seepage through the foundations of an earth dam While this seepage may be small it is important to provide an easy escape for it to prevent high uplift pressures under the dam This may be accomplished by providing a reverse filter in the downstream portion of the foundation which prevents the removal of fine material from the foundation by the seepage water

It is the usual practice to remove any unstable material such as soft clay fine sand silt vegetable matter and organic soil from the foundations of an earth dam. If the quantities of the unstable materials are too great to permit their removal special treatment must be given these materials to secure the necessary consolidation and stability One method of securing the required resistance to lateral movement especially where soft clays having low shear strength are encountered is to broaden the base of the dam and use very flat slopes at the upstream and downstream toes of the dam

Spillways All dams constructed of earth or rock and concrete dams over which it is undesirable to discharge water must have a spillway The only exceptions are dams built at offstream sites where the local runoff entering the reservoir is small in amount and can be safely stored in the reservoir without danger of overtopping the dam A spillway is a safety feature which provides for the release of water in excess of the reservoir capacity so that the dam and its foundations are protected from erosion and scour Ample spillway capacity is of particular importance for earth dams which would be severely damaged or even destroyed by flood waters passing over their tops

Spillways are of two general types the overflow type which is constructed as an integral part of the dam and the channel type which is an independent structure discharging through an open channel or tunnel depending on topography and cost Either type may be equipped with gates to control the discharge Various types of control structures have been used with the channel type of spillway including the simple overflow weir side channel overflow weir and the drop inlet or morning glory type where the water flows over a circular weir crest and drops directly into a tunnel

Unless the outlet of a spillway is sufficiently removed from the toe of the dam or erosion resistant bedrock exists at shallow depths some form of energy dissipator must be provided to protect the toe of the dam and the foundation from erosion by the high velocity water discharging from the spillway For an overflow spillway the energy dissipator may be a stilling basin a sloping apron downstream from the dam or a submerged "bucket The channel type of spillway usually has a stilling basin

Several types of gates may be used to regulate and control the discharge of spillways (Fig 8) Radial gates are comparatively low in cost and require only a small amount of power for their operation as they can be readily counterbalanced Drum gates are used to pass ice and drift over the tops of the gates They are relatively costly but they afford a wide unobstructed opening for the passage of logs and ice Operated by reservoir pressure they require no external power for their operation Vertical lift gates of the fixed wheel or roller type also are used frequently for spillway regulation They are comparatively costly and require considerable power for their operation Floating ring gates are used to control the discharge of the morning glory type of spillway Like the drum gate this type offers a minimum of interference to the passage of ice or drift over the gate and it requires no external power for its operation

Reservoir outlet works Reservoir outlet works control and regulate the release of water from the reservoir Outlets for concrete dams may consist of one or more conduits extending through the dam with control valves located either in a chamber or chambers in the dam or on the downstream end of the conduits or the outlets may be located in tunnels in the abutments of the dam Outlets for earth dams are usually located in one or more tunnels in the abutments of the dam but they may be placed in trenches excavated in the foundations or placed in conduits extending through the dam When conduits are located through or under an earth dam special precautions must be taken to prevent the leakage of water along the outside of the conduit

Outlet works usually have trashracks at the inlet end to prevent debris from clogging the outlets Usually some means is provided for closing off the upstream entrances to the outlet works by bulkheads head gates or stop logs so that the outlets may be unwatered for inspection and maintenance purposes

Outlet control gates Various types of gates and valves are used for controlling and regulating the release of water from reservoirs (Fig 9) These gates and valves must be free from excessive vibration and cavitation at any opening and at any head up to the maximum to which they may be subjected They must be capable of opening and closing under the maximum operating head and at the maximum velocity

Types of gates for regulating and controlling the release of water through reservoir outlets include high pressure slide gates radial or tainter gates needle valves of various kinds butterfly valves and cylinder valves or sleeve valves Emergency or guard gates are frequently used ahead of the operating gates at the larger and more important installations so that the service gates may be inspected and repaired without having to empty the reservoir

The slide gate consists of a movable leaf which slides on a stationary seat This type is used for low head service, it is not used generally for controlling the discharge of large quantities of water under high heads The high pressure gate is a type of slide gate of rugged design having a rectangular shaped leaf and bronze seats on both the movable leaf and the fixed body This type is used for regulating the discharge of water under operating heads up to 100 ft and for emergency or guard gates that may have to operate under unbalanced pressure only rarely under higher heads

Jet flow gates are a special form of slide gate in which the conduit immediately upstream from the leaf is contracted sharply at an angle of 45 degrees which causes the water to pass through the gate in the form of a free jet

Radial gates consist of a face plate supported by radial arms having bearings at their fixed ends to transmit the thrust to the supporting structure

Needle valves consist of a movable needle or plug enclosed in a stationary cylindrical body The needle may be opened or closed either hydraulically or mechanically The hollow jet valve is the latest form of needle valve used by the U S Bureau of Reclamation The discharge from this valve is in the form of a hollow or annular jet Butterfly valves consist of a circular disk or leaf mounted on a transverse shaft supported by two bearings, one on each end of the shaft The leaf is usually balanced hydraulically and only a small amount of power is required to operate the valve This type of valve is used mostly for shut off purposes for hydraulic turbines, pumps, and pipelines *See* VALVE

Penstocks. A penstock is a pipe that conveys water from a forebay, reservoir, or other source of supply to a turbine in a hydroelectric plant It is usually made of steel, but reinforced concrete and

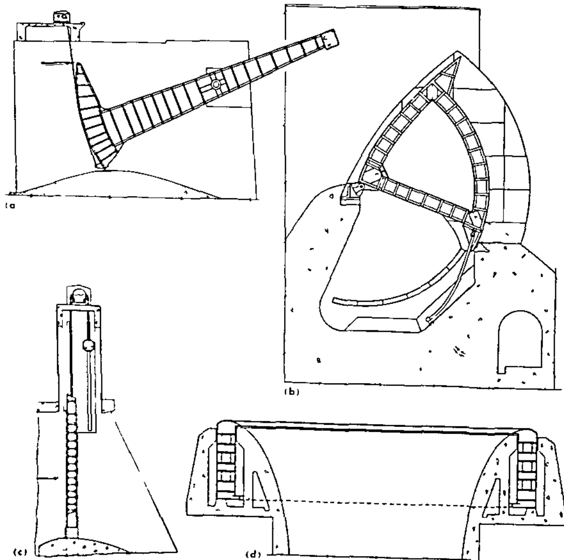


Fig 8 Typical spillway gates (*a*) Radial gate (*b*) Drum gate (*c*) Vertical lift gate (*d*) Ring gate (*U.S Bureau of Reclamation*)

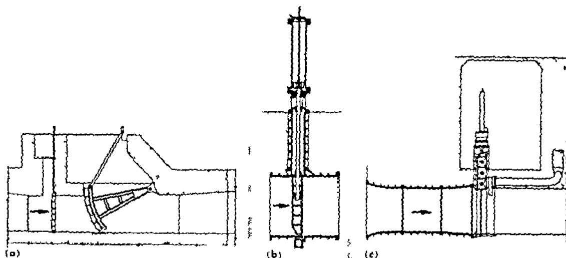


Fig 9 Typical outlet gates (a) High pressure slide gate (b) Jet flow gate (c) Radial gate in closed conduit (U.S Bureau of Reclamation)

woodstave pipe have been used for this purpose Pressure rise and speed regulation both have to be considered in the selection of the diameter of a penstock

Pressure rise or water hammer is a term applied to the pressure change that occurs when the rate of flow in a pipe or conduit is changed rapidly This intensity of the pressure change is proportional to the rate at which the velocity of flow is accelerated or decelerated Accurate determination of the pressure changes that occur in a penstock involves a number of factors many of them complex An important consideration, for example is the pressure rise that occurs in a penstock when the turbine wicket gates are closed after the loss of load *See* Canal Concrete Pipe flow [L N M]

Bibliography J Parmakian, *Waterhammer Analysis* 1955 U S Bureau of Reclamation *Trial Load Method of Analyzing Arch Dams* Boulder Canyon Project Final Reports, pt V Bull. 1 1938 H M Westergaard Water pressures on dams during earthquakes *Trans ASCE* 98 418–472 1933

Damping

A term broadly used to denote either the dissipation of energy in and the consequent decay of oscillations of all types or the extent of the dissipation and decay The energy losses arise from frictional (or analogous) forces which are unavoidable in any system or from the radiation of energy to space or to other systems. For sufficiently small oscillations the analogous forces are proportional to the velocity of the vibrating member and oppositely directed thereto the ratio of force to velocity is $-R$ the mechanical resistance For the role of damping in the case of forced oscillations where it is decisive for the frequency response *see* Forced oscillation Resonance (acoustics and mechanics) *See also* Harmonic motion Mechanical vibration Oscillation Vibration

Damped oscillations An undamped system of mass m and stiffness s oscillates at an angular frequency $\omega_0 = (s/m)^{1/2}$ The effect of a mechanical resistance R is twofold it produces a change in the frequency of oscillation, and it causes the oscillations to decay with time If u is one of the oscillating quantities (displacement, velocity acceleration) of amplitude A then

$$u = Ae^{-\alpha t} \cos \omega_d t \qquad (1)$$

in the damped case whereas in the undamped case

$$u = A \cos \omega_0 t \qquad (2)$$

The reciprocal time $1/\alpha$ in Eq (1) may be called the damping constant

In Eqs (1) and (2) the origin for the time t is chosen so that $t = 0$ when $u = A$ The damped angular frequency ω_d in Eq (1) is always less than ω_0 its value will be given later According to Eq (1) the amplitude of the oscillation decays exponentially the time

$$1/\alpha = 2m/R \qquad (3)$$

is that required for the amplitude to decrease to the fraction $1/e$ of its initial value

A common measure of the damping is the logarithmic decrement δ defined as the natural logarithm of the ratio of two successive maxima of the decaying sinusoid If T is the period of the oscillation then

$$\delta = \alpha T \qquad (4)$$

so that Eq (1) becomes

$$u = Ae^{-\delta t/T} \cos \omega_d t$$

Thus $1/\delta$ is the number of cycles required for the amplitude to decrease by the factor $1/e$ in the same way that $1/\alpha$ is the time required

The Q of a system is a measure of damping usually defined from energy considerations In the present case the stored energy is partly kinetic and partly potential when the displacement is a maximum the velocity is zero and the stored energy is wholly potential while at zero displacement the energy is wholly kinetic (*see* ENERGY) The Q is π times the ratio of peak energy stored to energy dissipated per cycle In the present example this reduces to

$$Q = \omega_0 m/R = \pi/\delta \qquad (5)$$

The damped frequency ω_d of Eq (1) is related to the undamped frequency ω_0 of Eq (2) by

$$(\omega_d/\omega_0)^2 = 1 - (1/2Q\ )$$

so that for high Q (lightly damped) systems it is only slightly less than ω_0

Overdamping critical damping If α in Eq (1) exceeds ω_0 then the system is not oscillatory and is said to be overdamped If the mass is displaced it returns to its equilibrium position without overshoot and the return is slower as the ratio α/ω_0 increases If $\alpha = \omega_0$ (that is $Q = ½$) the oscillator is critically damped In this case the motion is again nonoscillatory but the return to equilibrium is faster than for any overdamped case

Distributed systems An undamped one dimensional wave of frequency $\omega/2\pi$ propagated in the positive direction of x is represented by

$$u = A \cos \omega\ (t - x\ c) \qquad (6)$$

c being the velocity of the wave If the vibration is maintained at $x = 0$ at the value $u = A \cos \omega t$ then the damping manifests itself as an exponential decrease of amplitude with distance x Equation (6) is replaced by

$$u = Ae^{-\alpha' x} \cos\ [\omega(t - x/c)] \qquad (7)$$

The attenuation α' may depend on frequency If the medium is terminated the wave will be reflected from the ends and a system of standing waves will be set [illegible]

son [illegible]

or [illegible]

a system has a number of natural frequencies ω_n 2π at each of which it behaves like the lumped system of the previous sections The decay of a vibration is characterized by

$$Q = \omega_n/2\alpha' c = \pi/\alpha'\lambda_n \qquad (8)$$

where $\lambda_n = 2\pi c\ \omega_n$ is the wavelength *See* STANDING WAVE

Hysteresis damping At a given instant the elongation (strain) of a metal bar which is under periodic alternating stress is not determined exactly by the instantaneous value of the stress existing at that time (*see* STRESS AND STRAIN) For example the elongation is less at a given stress value when the stress is increasing than when it is decreasing This phenomenon which is known as mechanical hysteresis causes an undesirable energy loss A vibration problem of serious nature exists in the blades of jet engines and other steam and gas turbines The blade material itself exhibits a mechanical hysteresis damping which holds the vibrations in check When the stress is small the hysteresis damping is very small in all metals but it rises suddenly when the stress reaches a certain value Unfortunately in most metals the stress at which hysteresis damping becomes large and that at which the metal fails because of fatigue are very close together However a much higher hysteresis damping at safe stresses than that of ordinary steel is exhibited by certain alloys

Oscillating electrical circuits A simple series electrical circuit consisting of an inductance L resistance R and capacitance C is exactly analogous to the mechanical system described by Eqs (1) to (5) The inductance resistance and elastance ($1/C$) correspond to the mass mechanical resistance and stiffness respectively A distributed electrical circuit such as a section of transmission line or wave guide is analogous to a vibrating rod or disk *See* DYNAMICAL ANALOGIES

In the ordinary electrical oscillator the frequency is controlled by an electrical resonator (tank) lumped at the lower and distributed at the higher frequencies (*see* OSCILLATOR) Good frequency stability is associated with a high Q tank For frequencies of tens of megacycles per second and below mechanical resonators can be constructed which have a much higher Q than the equivalent electrical tanks Thus very stable electrical oscillators have mechanical resonators as their frequency determining elements Such an electromechanical system can operate only if there is some coupling between the electrical and mechanical aspects of the system

The coupling can be arranged in various ways In some materials such as quartz the constitutive relations involve the mechanical and electrical variables jointly thus for example an electric field may produce a strain in the absence of any stress Thus the coupling is inherent in the quartz itself (*see* PIEZOELECTRICITY) Quartz crystals are used for frequency control of oscillators in the range from kilocycles to perhaps 100 Mc the Q of a high frequency crystal may be several million Some low frequency oscillators are controlled by tuning forks having a Q of several hundred thousand the action being similar to that of the electric bell or buzzer *See* PIEZOELECTRIC CRYSTAL

The electrostatic motor generator effect provides the coupling in such mixed systems as the condenser microphone and electrostatic loudspeaker and the electromagnetic motor generator effect plays the same role in the dynamic microphone dynamic loudspeaker and in various electrical instruments

Reading of meters Galvanometers and other electrical indicating instruments are examples of damped electromechanical systems The free period depends on the moment of inertia of the rotating system (for example of the coil in a galvanometer) and on the stiffness of the suspension or

spring If an electrical input is suddenly applied the indicator will if the system is highly under damped overshoot its equilibrium value and then execute a damped sinusoidal oscillation about it while if the system is highly overdamped the indicator will approach its final reading sluggishly If the reading time is taken as the time required to reach the equilibrium value ±1% then the minimum reading time is obtained if the logarithmic decrement is 83% of the critical value (relative damping = 0 83) and is equal to 67% of the free period

In portable and switchboard instruments the damping is either viscous or magnetic or both Viscous damping is achieved with vanes attached to the movement which move in a narrow air filled space magnetic damping is an eddy-current effect (*see* EDDY CURRENT) The eddy currents are generated in the coil frame or in a metal plate moving between magnetic poles this latter arrangement is used in magnetically damped analytical balances

In the d Arsonval galvanometer the damping is largely due to the generator action of the moving coil and it can be adjusted by varying the external circuit resistance The same is true to a lesser extent of sensitive microammeters *See* GALVANOMETER *see also* ALTERNATING CURRENT CIRCUIT THEORY RESONANCE (ALTERNATING CURRENT CIRCUITS) TRANSIENT ELECTRIC VIBRATION DAMPING [M GR]

Bibliography F K Harris *Electrical Measurements*, 1952 L. E Kinsler and A R Frey *Fundamentals of Acoustics* 1950 N W McLachlan *Theory of Vibrations* 1951

Dark current

An ambiguous term used in connection with both gaseous discharge devices and photoelectric cells or tubes In gaseous conduction tubes it refers to the region of operation known as the Townsend discharge The name is derived from the fact that photons produced in the gas do not play an important part in the production of ionization The initial ionization arises from independent effects such as cosmic rays radioactivity thermionic emission or similar processes When applied to photoelectric devices, the term applies to background current This is current which may be present as the result of thermionic emission or other effects when there is no light incident on the photosensitive cathode *See* ELECTRICAL CONDUCTION IN GASES TOWNSEND DISCHARGE [C H MI]

Darter

Any of about 100 species of North American fresh water fishes of the subfamily Etheostomatinae family Percidae The darters are all slender small bottom dwelling fishes The majority live in swift streams with rocky bottoms but a few occur in lakes and sluggish streams Their swimbladder is degenerate a modification to their bottom dwelling habits Many of the darters are brilliantly colored and are among the most highly colored of all fresh



Johnny darter, *Etheostoma nigrum* (*From* E L *Palmer Fieldbook of Natural History McGraw-Hill 1949*)

water fishes They feed upon insects and zooplankton *See* PERCIFORMES. [J D B]

Dasheen

The plant *Colocasia esculenta* and the variety *antiquorum* which is taro are among the few edible members of the aroid family (Araceae) These plants are natives of southeastern Asia and Malaysia where they supply the people with their most



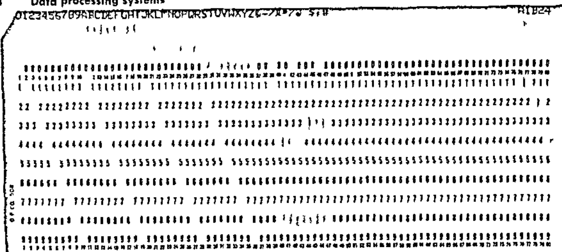
Colocasia esculenta (*Caladium esculentum*) (*L H Bailey The Standard Cyclopedia of Horticulture vol 1 Macmillan 1937*)

important food The edible corms (underground stems) support a cluster of large leaves 4–6 ft long A main dish in the Polynesian menu is poi a thin pasty gruel of taro starch often fermented which is frequently formed into cakes for baking or toasting The corms are baked or boiled to eliminate the irritating substance present in the cells of the raw corm *See* ARALES [P D S]

Data processing systems

Mechanical electromechanical, or electronic machines for transforming information given to the system in unorganized form into a suitably organized presentation according to some criterion and according to some orderly preplanned procedure The term data processing system is also applied to the scheme, or procedure which the machines employ in processing information

Typical business applications of data processing systems are for payrolls maintenance of accounts, billing report preparation stock recording and inventory control and sales analysis In science and engineering data processing systems are used in data reduction preparation of mathematical and statistical tables recording and maintenance of experimental data and revision and correction of libraries of routines *See* DATA REDUCTION



Fg 1 Punched card A1B24 coded in columns 76–80

Data processing systems may be classified and described according to the degree in which manual intervention is required to control and execute the processes the system is meant to carry out Classified in this way the two principal types are punched card systems and magnetic tape systems This classification also characterizes the medium for bulk storage of information and the chief means of input and output for the respective systems

Punched card systems These usually comprise one or more key punches sorters collators reproducers calculating machines interpreters and printers

Punched cards (Fig 1) are initially produced by manual operation of a key punch having a typewriterlike keyboard Key punches are constructed to code information according to one of the card codes of Fig 1 In data preparation coding occurs in two or more steps For example in a stock inventory system the code A1B24 may be assigned to 10-mil steel wire Recording in a card of an amount of wire in inventory would include a second encodement (of A1B24) into a pattern of punched holes

Reproducers In some applications cards are marked by hand with an electrically conductive material These must be converted to an equivalent set of punched cards by a reproducing machine because all other card machines read holes either by wire brushes which make electrical contact through the holes with a revolving metal cylinder or by photoelectric means The reproducing machine is controlled by electrical impulses originating in the reading process A control panel has provisions for interchanging the connecting wires and thereby changing the card punching operation of the machine The control panel for an 80-column card includes 80 exit hubs from the card reading system and 80 entry hubs to the actuating system of the punches By manual insertion of wires the columns 1–10 of the marked card for example may be punched into positions 80 90 of the card copy Reproducing machines can also be used to punch new cards from previously punched cards These are used principally to create several copies of a deck for simultaneous processing on other machines for suppression or rearrangement of card columns through panel wiring and for producing fresh unworn decks

Sorters Capable of processing up to 2000 cards per minute sorters are used to arrange cards in order according to a key or field For example a file of payroll cards may be sorted into alphabetical order by last name where the name is the key Sorters sense holes of a single column only and act by disposing each card into one of 13 pockets corresponding to each of the card rows or to a blank column Information occupying more than a column is arranged by iteratively sorting on a sequence of columnar positions

Collators These machines which have two sets of card feeds and sensing brushes are used to match or merge together two files of cards according to some criterion of identity, such as name or part number Merging operations are preceded by sorting for files not previously arranged according to the key

Both sorters and collators may be used for selecting data of a described class (for example a 9 punch in column 10) Sorters dispose cards into pockets by a brush sense setting collators match one file of cards against a second file A wide variety of classificatory, matching and selecting processes can be accomplished on either machine or a combination of both

Printed representation of card information may be obtained on cards themselves by interpreters for ease of human reading Control panel wiring makes it possible to suppress or rearrange the information printed across the top of a card in an order not necessarily the same as that of the holes

Printers Printers including accounting machines or tabulators are the principal means of summarizing and displaying processed information for the human user Most printers employ panel wiring to allow variation of the format of the output information Most printers employ some direct means of printing the output data on paper and

operate at speeds up to 1000 lines of 120 characters or more each per minute Some printers employ photography of characters displayed on the face of a cathode-ray tube and xerographic printing *See* PHOTOCOPYING PROCESSES STORAGE TUBE

Magnetic tape processing systems These differ from large scale scientific computers in application rather than in the essential character of the equipment used For large-scale computers *see* DIGITAL COMPUTER However data processors often have larger bulk storage capacity and more flexible input and output equipment than their scientific counterparts Considerably slower speeds of operation and use of coded decimal and alphabetic character representation (in contrast to use of the pure binary number system) are further distinguishing properties

Oxide coated plastic tape ½ in wide on reels of 1500–2500 ft is generally used although some systems use a bronze phosphor alloy tape Information is recorded in the form of magnetic spots with a density of 100–500 per in and in 7 or 8 to 30 channels (on 3 in tape) Tape reading and writing speeds vary from 50 to 150 in /sec Using a six bit binary code reading and writing rates of 15 000–75 000 alphabetic or numeric characters per second are practical (Fig 2)

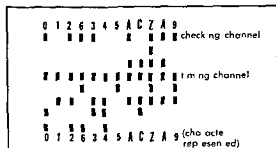


Fig 2 Strip of 8 channel magnetic tape Two channels are used for timing and checking The other six are used to represent an alphanumeric character in a 6 bit excess-3 code

Magnetic tape read and write circuitry is connected through amplifiers to the central processor by way of special storage registers called buffers These serve to match the speed of the computer with that of the tape and thereby enable the computer to process information simultaneously with the input and output operations

The central processor is essentially a decimal digital computer with magnetic drum electrostatic magnetic core or acoustical delay line high speed memory serial mode arithmetic circuitry and control circuitry *See* STORAGE DEVICES

Unlike punched card systems magnetic tape systems operate automatically under a program which is itself coded and stored on tape and executed from the internal memory of the central processor Sorting merging and arithmetical operations are accomplished by programmed routines in a quite analogous to that in scientific computation

Input to tape may be direct via keyboard or indirect from punched cards using special card-to-tape converting machines Output may be direct from tape to printer or indirect from tape to punched cards to printer via a converter Auxiliary equipment such as printers punches and special sorters may be under the programmed control of the central processor in which case such equipment is said to be on line otherwise it is off line

For applications requiring nonsequential or random interrogation of a file tape processing systems may be augmented by the use of disk storage devices These consist of a set of magnetizable rotating disks which except for magnetic reading and writing means functions in a way analogous to a multirecord phonograph On 50 disks up to 5 000 000 characters may be stored with an average access time to any character of about ½ sec or less Tape access on the other hand averages at least 60 sec unless the wanted information is presorted into proper sequence for interrogation

Punched paper tape is frequently employed as input to a processing system as well as to digital computers in general and may use any of the standard 5 to 8 hole systems of coding Both mechanical (up to 60 characters per sec) and photoelectric (up to 200 characters per sec) readers are commonly used

An advantage of paper tape is the possibility of long range communication (for example by teletype) between various inputs and a central processor Systems using paper tape are however limited in speed [R J N]

Bibliography C C Gotlieb and J N P Hume *High Speed Data Processing* 1958 Review of input and output equipment used in computing systems *Proc Joint AIEE IRE ACM Computer Conference* 1952

Data reduction

The transformation of information usually empirically or experimentally derived into corrected ordered and simplified form

The term data reduction generally refers to operations on either numerical or alphabetical information digitally represented or to operations which yield digital information from empirical observations or instrument readings (*see* DIGITAL COMPUTER) In the latter case data reduction also implies conversion from analog to digital form either by human reading and digital symbolization or by mechanical means *See* ANALOG TO DIGITAL CONVERTER

Data reduction is used to prepare data in a form suitable for scientific computation statistical analysis and control of industrial processes and operations and for data processing in business applications Examples are the preparation of data from test runs in missile development win experiments industrial product sampl ings of sensing instruments in

In applications where the raw data are already digital data reduction may consist simply of such operations as editing scaling coding sorting collating and tabular summarization

More typically the data reduction process is applied to readings or measurements involving random errors These are the indeterminate errors inherent in the process of assigning values to observational quantities In such cases before data may be coded and summarized as above outlined the most probable value of a quantity must be *determined* Provided the errors are normally distributed the most probable (or central) value of a set of measurements is given by the arithmetic mean or in the more general case by the weighted mean *See* STATISTICS

Data reduction may also involve operations of smoothing and interpolation because the results of observations and measurements are always given as a discrete set of numbers while the phenomenon being studied may be continuous in nature

In a smoothing problem a function is empirically given (for example positions of a body as a function of time) as a collection of points $(t_1 x_1)$ $(t_2 x_2)$ $(t_n x_n)$ where the values of the variables perhaps both independent and dependent are inaccurate A common procedure is to fit an nth (commonly second) order parabola by least squares to the data points thus obtaining a representation that will satisfy as nearly as possible all of the given pairs but perhaps none of them exactly

In interpolation a function is known in tabular form The problem is to determine values between the tabulated points *See* INTERPOLATION

Any of the above mentioned procedures may be carried out on a digital computer or built into a process control system or procedure [R J N]

Data transmission

The conveying of information usually in rigid format from one location to another The two locations may be close enough for direct mechanical connection or so widely separated as to require a variety of communications techniques such as radio relaying The form of the information usually differs from the messages handled by such communication channels as telephone telegraph or television and is generally quantized into discrete steps as distinguished from the analog *representation* used in some telemetering circuits The data are often transmitted rapidly and electronic techniques are used

Many principles of data transmission are descended from long usage in the fields of telephony telegraphy facsimile and accounting machinery However the over all design of systems for data transmission is new Clerical accounting methods and procedures are being developed rapidly to employ electronic data transmission

Data transmission is generally between a point or points of activity and a central location where records are kept and the data are processed Increasingly both record keeping and data processing are done automatically with electronic equipment The point of activity may be a sales office production line shipping platform cash register ticket window or laboratory test equipment

The fundamental components of an over all data transmission system are data input output terminal equipment and communication facilities

Input output equipment Fixed data (recurrent items such as part numbers) are frequently prepunched in tape or cards so that they can be 'read' by a machine unit Variable data (quantity price and so forth) which are unique to each transaction are entered by means of a keyboard which constitutes the man to machine linkage The electronic output from these is in machine sensible language that is it can be understood by business machines

The form size shape and construction of the keyboard vary widely being dependent upon its intended use The same is true of the readers For applications where usage is light a simple device may be provided One experimental model includes a 10-button keyboard on top and a spring powered card reader on its side This particular unit is designed for transmission via telephone connections For handling higher volumes (bulk data transmission) fully automatic high speed readers of paper tape cards or magnetic tape are generally used At the receiving locations a card or tape punch or magnetic tape recording device may be used to record the received data in machine sensible form Alternatively the data may be recorded in human sensible form by a printer operated either on line or from cards or tapes which were recorded on line

Data terminal equipment The electrical bits (binary digits) in yes no pulse form are accepted by terminal equipment that conditions them for transmission over the communications channel generally as tones This equipment or unit is sometimes referred to as a subset because its function *for data transmission is comparable to the function* of the telephone subset (subscriber set) for voice transmission that is it converts the electronic data bits into suitable form (tones) for transmission at the sending end and vice versa at the receiving end

An essential component of the subset is a modulator for transmitting and a demodulator for receiving The designer s choice of the type of modulation (AM FM or phase) used in a particular system is based on such considerations as speed requirements and economically justifiable complexity Transistors and other miniaturizing and power saving components are used Power may be supplied *centrally* as it is in telephone systems from a central office battery Various multiplexing techniques are used For example the equipment may transmit simultaneous (parallel) frequencies on a 3-out of 12 basis handling up to about 200 bits/sec over a voice frequency channel

From the receiving subset the data may be fed directly into a data processing unit such as a

computer, or it can operate a card or tape unit to record the data on a storage medium in machine-sensible form The data processing equipment and recording devices such as card punches, tape units, or high speed printers are usually considered part of the business machine equipment

Communications facilities. The over all circuit that connects the two machines involved in a given data communication is generally composed of both channels and switches The channels may be part of a transmission system (having other uses) on wire lines, coaxial cable, or microwave radio The switches may be a telephone central office or the equivalent Special purpose systems such as the SAGE (semiautomatic ground environment) system for air defense, may permanently connect the data processing unit to remote data input output units

Channels used for data transmission are generally of the same type as those used for voice communication However, some types of high speed data transmission involve requirements which are more exacting than those for voice transmission for example, more uniform amplitude and phase response, as well as lower impulse noise The ear is more tolerant of amplitude and phase (delay) distortion and of impulse noise than are machines Hence, some data systems require specially equalized and otherwise treated channels Tests and usage experiences have shown that most regular telephone circuits are adequate for most cases of data transmission, in the case of those which are not it is generally possible to make rather simple compensations in the terminal equipment to achieve satisfactory over all performance

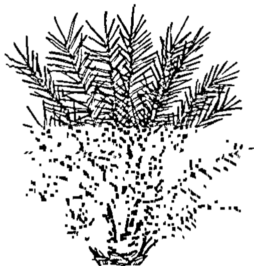
The bandwidth of the channel imposes an upper limitation on the speed of transmission The limitation may vary with the refinement built into the data terminal equipment Typical low cost data terminals may operate in the range of 75–200 bits/sec over voice frequency facilities More complex and expensive terminals may permit operation at speeds of 600–3000 bits/sec over the same facilities For a particular system the most economic arrangement is a balance between line and terminal costs, the most expensive terminals being justified where there are large volumes of data to be transmitted over long distances

The error performance and the controlling of the error rate with typical circuits is comparable to that experienced in the use of magnetic tape in computer applications Just as the magnetic tape is a weak link in the computer system and must be protected by the inclusion of redundancy, so is it customary to transmit redundancy over communication channels to permit the detection and correction of any errors which might otherwise be introduced The most effective transmission error control systems are designed so that the redundancy is time separated from the information bits to which it is related See TELETYPEWRITER EXCHANGE (TWX) SERVICE

[V N V]

Date

The evergreen date palm, *Phoenix dactylifera*, is dioecious (each sex on different plant) Pollen, borne only on male palms, must be transferred by hand to inflorescences on female tree to induce fruit production (*see* INFLORESCENCE; REPRODUCTION, PLANT) Date fruits are dry, semidry, or soft,



Young date palm Trees may reach 100 ft in height (From L H Bailey, *The Standard Cyclopedia of Horticulture*, vol 1, Macmillan, 1937)

depending on the moisture content at maturity of the particular variety Soft types tend to spoil or sour more readily than dry types Most varieties contain about 65% sugar as glucose or fructose One variety, Deglet Noor, grown in California contains only sucrose sugar Trees are propagated by offshoots or suckers and are grown mainly in the semiarid and desert regions of Asia and Africa Commercial date culture in the United States limited to areas of high heat (100–120°F) and low atmospheric humidity, such as the low elevation deserts of California and a small part of Arizona results in fruit having an annual value of approximately $2,050,000 *See* FRUIT (BOTANY), FRUIT (TREE), PALMALES, *see also* FOOD ENGINEERING

[C A S]

Dating methods

Methods and techniques used in archeology biology, and geology to fix dates, assign periods of time, and determine age The uranium lead, rubidium strontium, potassium argon, and carbon 14 methods have yielded reliable results on suitable samples and have permitted the construction of absolute geologic history in many areas of the world *See* RADIOCARBON DATING, ROCK (AGE DETERMINATION), *see also* EARTH (AGE OF), GEOLOGIC TIME SCALE, LEAD ISOTOPES, GEOCHEMISTRY OF, METEORITE

For descriptions of other methods and techniques used to establish the sequence of events the succession of strata and relative chronologies *see* ARCHEOLOGY CHEMICAL DATING CLIMATIC CHANGE DENDROCHRONOLOGY GEOLOGY, INDEX FOSSIL MARINE SEDIMENTS PALEOBIOCHEMISTRY, PALEOBOTANY PALEONTOLOGY PALYNOLOGY POSTGLACIAL VEGETATION AND CLIMATE ROCK MAGNETISM STRATIGRAPHY TREE RING HYDROLOGY, UNCONFORMITY VARVE *See also* GEOCHRONOMETRY

Datolite

A mineral nesosilicate composition $CaBSiO_4(OH)$, crystallizing in the monoclinic system It usually occurs in crystals showing many faces and having an equidimensional habit It may also be fine granular or compact and massive Hardness is 5 5½ on Mohs scale specific gravity is 2 8 3 0 The luster is vitreous the crystals colorless or white with a greenish tinge See SILICATE MINERALS Datolite is a secondary mineral found in cracks and cavities in basaltic lavas or similar rocks associated with zeolites apophyllite prehnite and calcite It is found in the Harz Mountains Germany Bologna Italy and Arendal Norway In the United States fine crystals have come from Westfield Massachusetts Bergen Hill New Jersey and various places in Connecticut In Michigan in the Lake Superior copper district datolite occurs in fine grained porcelainlike masses which may be coppery red due to inclusions of native copper [C S HU]

Day

A unit of time equal to the period of rotation of Earth Different sorts of day are distinguished according to how the period of rotation is reckoned with respect to one or another direction in space

Solar day The apparent solar day is the interval between any two successive meridian transits of the Sun It varies through the year reaching about 24 hr and 30 sec of ordinary clock time in December and about 23 hr 59 min and 39 sec in September

The mean solar day is the interval between any two successive meridian transits of an imagined point in the sky that moves along the celestial equator with a uniform motion equal to the average rate of motion of the Sun along the ecliptic Ordinary clocks are regulated to advance 24 hr during a mean solar day

Sidereal day The sidereal day is the interval between any two successive meridian transits of the vernal equinox Similarly as for the solar day a distinction is made between the apparent sidereal day and the mean sidereal day which however differ at most by a small fraction of a second A mean sidereal day comprises 23 hr 56 min and 4 09054 sec of a mean solar day

The period of rotation of Earth with respect to a fixed direction in space is 0 0084 sec longer than a sidereal day No special name has been given to this kind of day and although of theoretical interest it is not used in practice

Variations in duration The mean solar day the sidereal day and the day mentioned in the preceding paragraph all vary together in consequence of variations in the speed of rotation of Earth which are of three sorts seasonal, irregular, and secular The seasonal variations are probably caused at least in part by the action of winds and tides the effect is to make the day about 0 001 sec longer in March than in July, and is nearly repetitive from year to year The irregular variations are probably the result of interactions between motions in the core of Earth and the outer layers, the effect is to cause more or less abrupt changes of several thousandths of a second in the length of the day which persist for some years The secular variation is the result of tidal friction mainly in shallow seas which causes the duration of the day to increase about 0 001 sec in a century *See* TIME [G M C]

Bibliography H Jeffreys *The Earth, Its Origin History and Physical Constitution* 3d ed 1952

De Broglie wavelength

The wavelength $\lambda = h/p$ associated with a beam of particles (or with a single particle) of momentum p $h = 6\ 61 \times 10^{-27}$ erg sec is Planck's constant The same formula gives the momentum of an individual photon associated with a light wave of wavelength λ This formula along with the profound proposition that all matter has wavelike properties was first put forth by Louis de Broglie (1924) and is fundamental to the modern theory of matter and its interaction with electromagnetic radiation *See* QUANTUM MECHANICS, QUANTUM THEORY NONRELATIVISTIC [E C]

Dead reckoning

A form of navigation that determines position of a craft by projecting from a previous to a new position on the basis of assumed distance and direction moved The name probably stems from the early practice of determining speed by throwing overboard a buoyant object called a Dutchman's log and noting the time needed for a known length of the vessel to pass the floating object or attaching a line to the object (when the whole device became known as a chip log) and noting the amount of line paid out in a given time In either case, the floating object was assumed to remain dead in the water providing an indication of speed through the water The reckoning of future positions of the vessel by means of this speed was known as dead reckoning

Parameters The parameters of dead reckoning are direction of motion and distance traveled Several directions are involved in the motion of a craft as shown exaggerated in Fig 1 The intended direction of travel is called the course The direction steered which may differ somewhat because of anticipated offset due to wind or current is also called course (or course steered) by the marine navigator and heading by the air navigator The direction actually made good between two points is called course made good Air navigators often refer to this direction as the track although this expression is used also to refer to the path fol

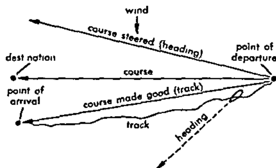


Fig 1 Directions involved in the motion of a craft

lowed or intended to be followed The direction in which the craft is pointed at any moment is called heading by both air and marine navigators

A compass is used to indicate direction Distance is usually determined indirectly by measurement of speed and time, but it may be measured directly

Uncertainties, offset, and errors Air and land navigators, and some marine navigators use the best estimate of direction and distance in their dead reckoning Many marine navigators, however, prefer to use course steered and estimated speed through the water (without allowance for the effect of wind) for their dead reckoning considering positions determined by allowance for estimated effects of wind and current as estimated positions

The uncertainty of dead reckoning positions however determined increases with time and perhaps also with distance traveled From time to time an independent determination of position is made by celestial navigation or piloting When a reliable position called a fix is so obtained a new dead reckoning is started from this point

The average combined effect of wind, current and other sources of error since the last fix can be determined by comparison of a fix with the position the craft would have occupied had there been no disturbing force It is for this reason, in part that many marine navigators prefer to use course steered and speed at which the vessel is propelled by its primary source of power for their dead reckoning For convenience, they consider the entire offset effect as the result of current The direction from the dead reckoning position to a fix at the same time is called the set of the current The distance between these two positions divided by the time since the last fix is called the drift (speed) of the current Air navigators consider the entire offset effect as the result of wind, determined by comparison of a no wind position, sometimes called an air position with a fix at the same time

Whatever other forms of navigation may be employed dead reckoning is generally considered essential to safe navigation It provides a continuous position which is valuable in the evaluation of all other navigational information received It is always available in some form

Determination of position. Two methods, plotting and computing and a mechanical computer application, are most used for determination of position

Plotting The determination of position by dead reckoning is commonly performed by plotting on the chart or plotting sheet Marine navigators usually measure direction by means of parallel rulers (a device designed to move parallel to itself) or a drafting machine (which combines parallel motion with direction indication) Dividers are usually used for measuring distance Air navigators usually use some form of plotter, which consists of a protractor with an attached straightedge The straightedge usually carries a scale for measuring distance, but dividers may be used for this purpose One type of plotter, in wide use among air navigators is shown in Fig 2

Computing Before reliable charts and plotting sheets became generally available dead reckoning was usually performed by computation Various forms of computation collectively known as the sailings, were developed These are seldom used by modern navigators except in small craft where chart work is difficult, or for computing direction and distance between widely separated points

Mechanical computor method In many larger ships and aircraft the dead reckoning is performed mechanically by a device that receives inputs of direction and speed and automatically computes dead reckoning positions, which are displayed continuously on dials or traced on a chart or plotting sheet

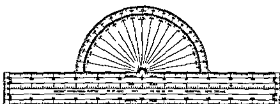


Fig 2 A plotter in wide use among air navigators

Refinements of direction measurement. Accurate dead reckoning requires reliable measurement of direction and distance Directions are expressed as angular distance from a reference direction usually north for courses True north is the direction of the North Pole, magnetic north is the direction north along a magnetic meridian, compass north is the direction north as indicated by a magnetic compass, and gyro north is the direction north as indicated by a gyrocompass For another reference direction, grid north, *see* POLAR NAVIGATION Courses are customarily stated in whole degrees, using three figures from 000° at north clockwise to 360°

Courses have not always been so stated, nor this system universally used now The points (and half and quarter po monly used for many centuries An of the education of former tion in boxing the compass or r points of the compass

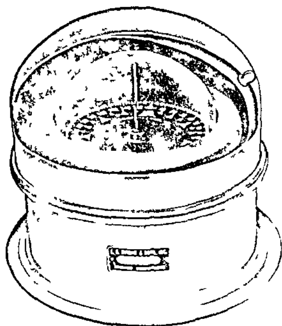


Fig 3 A typical modern marine magnetic compass in a navy type binnacle

Magnetic compass For many centuries the primary direction instrument has been the magnetic compass The first such compass probably consisted of a magnetic needle freely floated in water by means of a straw This was replaced by the dry compass and this in turn by the modern liquid compass in which the north seeking magnetic element is placed in a compass bowl completely filled with a liquid which will not freeze at the temperatures in which the compass might reasonably be expected to operate

A typical modern marine magnetic compass is shown in Fig 3 The directional element consisting essentially of several bundles of slender magnetized rods is mounted by means of an almost frictionless bearing so that the compass will be responsive to weak magnetic fields

Attached to the magnetic element is an annulus of lightweight material on which compass graduations are placed This is the compass card which remains aligned with the earth s magnetic field as the vessel turns Heading is indicated by the compass graduation aligned with a lubber s line in the longitudinal axis of the craft generally on the forward side of the compass bowl aboard ship and the after side in aircraft and land vehicles Bearings can be measured by some compasses by means of a suitable attachment provided with sighting vanes

Magnetic compasses are subject to compass error, the difference between true north and compass north This error is the algebraic sum of two components One of these called variation by the navigator and magnetic declination by the magnetician is the angle between magnetic and geographic meridians It is related to the magnetic field of the earth and is subject to a small daily or diurnal change, and to a slow progressive secular change It is affected also by magnetic storms

The second component is deviation the angle between the magnetic meridian and the axis of the compass card It is the result of local magnetic influences particularly those within the craft Permanent magnetism in metal of the craft and transient magnetism induced there by the earth's magnetic field are important contributors Direct currents in electrical wiring also influence the compass *See* COMPASS MAGNETIC

The various disturbing influences can be largely neutralized by establishing additional magnetic fields of equal strength but opposite polarity This process is called compass adjustment, or sometimes compass compensation especially with respect to aircraft compasses In the U S Navy the latter expression is used to indicate the process of neutralizing the effects that degaussing currents exert on a magnetic compass These currents may be used to change a ship's magnetic characteristics as a protection against magnetic mines

In some compass installations particularly in aircraft the magnetic compass is installed in a position relatively free from magnetic disturbances and its indications are transmitted electrically to locations throughout the craft Such an instrument is called a remote-indicating compass

Compass error is determined by comparing it with a compass having a known error or by azimuth of a celestial body or bearing of a landmark

When deviation is reduced to a minimum the residual deviation on a number of headings is recorded on a deviation table or deviation card kept near the compass A typical deviation card used in aircraft is shown in Fig 4

The directive force of the compass is the horizontal component of the earth s magnetic field at the compass This becomes progressively weaker as the magnetic poles are approached Within a

#2 MASTER COMPASS
SWUNG 7 24 58 BY KB

TO FLY	STEER	TO FLY	STEER
N	359	180	181
15	15	195	196
30	30	210	211
45	44	225	226
60	60	240	241
75	75	255	256
90	90	270	271
105	105	285	286
120	121	300	301
135	135	315	316
150	151	330	330
165	166	345	345

Fig 4 A typical aeronautical deviation card

few hundred miles of these poles the magnetic compass is unreliable See GEOMAGNETISM

Gyrocompass In addition to several magnetic compasses nearly all naval vessels and ocean liners are equipped with one or more north seeking gyrocompasses This instrument has the disadvantage of requiring a source of electrical power but it is not subject to variation or deviation It may have a small gyro error but in most modern installations this is not large enough to be significant except in high latitudes The compass tends to align its rotational axis with that of the earth but since it is constrained to remain in the horizontal its directive force decreases as the horizontal becomes more nearly perpendicular to the earth s axis Beyond some latitude generally about 65° the gyro error becomes large and erratic and the instrument requires frequent checking Near the geographical poles the gyrocompass becomes unreliable

Great progress has been made in the development of better gyroscopes As a result gyrocompas es have become smaller lighter more accurate and practical for smaller vessels Some work has even been done to develop one for aircraft

Directional gyro A directional gyro used in some aircraft compass systems tends to maintain a fixed position in a great circle plane The better instruments are automatically precessed to allow for rotation of the earth so that a craft following their indications tends to follow a great circle on the earth Since the instrument does not seek anything it is subject to a slow cumulative error called wander This is allowed for or corrected at intervals by comparison with a directional reference generally a magnetic compass or the azimuth of a celestial body See AIRCRAFT COMPASS SYSTEM

Several types of devices provide directional guidance in polar regions See POLAR NAVIGATION

Distance or speed measure Distance is generally stated by navigators in units of nautical miles usually to integral miles by air navigators and to tenths of a mile by marine navigators For this and other linear units of navigation *see* NAVIGATION

Aboard ship distance or speed is measured by means of a log or by an engine revolution counter The primitive Dutchman s log and chip log were mentioned earlier The latter is of particular interest because a series of knots was tied in the log line and the number of knots which passed through the seaman s hands in a given time interval represented the speed of the vessel in nautical miles per hour This is the origin of the name knot applied to the unit of speed still in common use

About the middle of the seventeenth century mechanical logs appeared These had rotators towed through the water first with the indicator also in the water but later at the taffrail giving the name to the taffrail log widely used for many years

The pitometer log now in wide use uses a Pitot static tube The Forbes log uses a small rotor in a t ibe projecting below the bottom of the vessel electromagnetic log has a sensing element which produces a voltage directly proportional to speed through the water Accurate measured miles are marked out at various places on the beach to provide means for calibrating or checking logs

In aircraft speed through the air is measured by means of an air speed indicator or Mach meter The latter provides an indication of speed in units of the speed of sound which varies with density of the atmosphere For measurement of air speed a Pitot static tube is generally used with a suitable registering device Distance is determined indirectly from speed and time Distance and indirectly speed may also be determined by means of two fixes the length of the line between them being the distance See AIR VELOCITY MEASUREMENT

Dead reckoning systems Two sophisticated system of performing the entire dead reckoning process have been developed One of these determines course and speed of an aircraft over the surface by measuring the Doppler shift of echoes of pulse modulated beams of radio energy directed downward at an angle The other system employed both in aircraft and in ships uses accelerometers and gyroscopes with a good heading reference to determine accelerations in various directions These are converted by integration to components of speed and distance in each direction Both systems include computers to provide continuous indication of position See CELESTIAL NAVIGATION INERTIAL GUIDANCE SYSTEM NAVIGATION SYSTEMS ELECTRONIC PILOTING [A B M]

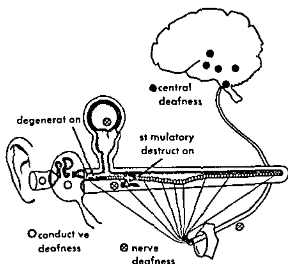
Bibliography N Bowditch *American Practical Navigator* U S Navy Hydrographic Office H O 9 1958 J C Hill II T F Utegaard and G Riordon *Dutton s Navigation and Piloting* 1958 U S Navy Hydrographic Office *Air Navigation* H O 216 1955

Deafness

Temporary or permanent impairment of hearing An individual s limits of hearing are represented graphically as an audiogram which shows relative sensitivity to different frequencies of sound as compared with normal hearing Temporary deafness may be the result of physical injury or illness or of auditory fatigue that is the decrease in sensitivity after exposure to sound Permanent deafness from sound exposure is called acoustic trauma or stimulation deafness Stimulation deafness caused by specific sound frequencies may be restricted to a frequency range somewhat higher than the stimulating frequency Permanent deafness may also be the result of injury disease or developmental anomalies either early in life or during aging Deafness is rarely complete but may involve complete insensitivity at some frequencies In partial deafness the impairment is usually greater at high frequencies than at low

The classification of deafness is based on its theoretically defined origin in the middle ear cochlea auditory nerve complex or the brain

uction or injury of the eardrum or im



Types of deafness with illustration of cochlear degeneration and stimulatory destruction of the organ of Corti Degeneration of the organ of Corti near the base of the cochlea is characteristic of old age The cochlea is shown here as uncoiled

sponse of the bones or membranes of the middle ear results in conduction deafness which is always partial and usually involves general loss of sensitivity for all sound frequencies Nerve deafness may arise from some malfunction of the cochlea the cochlear canal the basilar membrane the organ of Corti or the auditory nerve as a result of biochemical physiological or anatomical disturbances or defects Stimulation deafness involves changes in the chemical interchange between the cochlear canal and other canals of the cochlea as well as possible destruction of the receptors and nerve fibers in the organ of Corti

Central deafness is produced by defects or disturbances in brain function such as tumors Anything that can affect neural integration can bring about defects in hearing

Many phenomena of deafness are significant in their diagnosis Tinnitus or ringing in the ear may occur with overexposure to sound progressive deterioration of hearing during aging or disease or irritative disease of the middle ear Diplacusis is a difference in the pitch perceptions of the two ears when stimulated by the same sound frequency Although a certain amount of diplacusis is characteristic of normal hearing an exaggerated condition is usually the result of an abnormality of the inner ear Recruitment is the apparent increase in loudness of tones to the partially nerve deaf or cochlear-deaf ear over normal loudness In such deafness perception of loudness becomes more and more normal as the intensity of stimuli is increased Thus once a tone is over the threshold of the nerve-deaf ear and continues to increase in intensity its loudness increases more rapidly than it would for the normal ear

Deafness is widespread especially in industrial populations It most often results from sound overexposure over long periods of time and the physiological deterioration of aging Progressive hearing loss for high frequencies is characteristic of the aging process On the average people of 30 years show decreased sensitivity for frequencies over 4000 cycles per second (cps), those of 40 over 2000 cps those of 50 over 1000 cps and those of 60 over 250 cps In old age frequencies over 6000–8000 cps may not be heard at all

Nearly complete or complete deafness involves severe stress effects on the individual as well as general psychological impairment related to perceptual restriction Therapy for deafness involves treatment of the symptomatic hearing loss if possible and general rehabilitative preparation of the individual for effective social communication and work *See* AUDIOMETRY HEARING [K U S]

Bibliography S S Stevens (ed) *Handbook of Experimental Psychology* 1951 I J Hirsh *The Measurement of Hearing* 1952

Death

The cessation of life Death may involve the organism as a whole (somatic death) or may be confined to cells and masses of cells within the organism (necrosis and necrobiosis)

Somatic death The death of the body in its entirety is usually characterized clinically by the cessation of cardiac activity circulation and respiration Certain cells and tissues in the body may remain alive for limited periods of time after somatic death but as the temperature of the body falls metabolic activity ceases

Somatic death is characterized by a number of irreversible changes which are of medicolegal importance especially for purposes of estimating the time of death These include such phenomena as rigor mortis livor mortis algor mortis autolysis and putrefaction

Rigor mortis Rigor mortis is the muscular rigidity which begins about 4–12 hours after death in most instances depending on the temperature presence of infection and certain other factors It passes off within 12 36 hours Its first appearance in skeletal muscle is usually seen about the head and neck but eventually all the muscles are involved The actual cause of rigor is obscure but its onset coincides with the swelling and precipitation of certain proteins as the tissue becomes acid With a further increase in acidity the proteins solubilize and the rigor ceases

Livor mortis Livor mortis is the red discoloration of the dependent portions of the body as a result of the settling of blood Livor is usually apparent 1–2 hours after death and after 10–12 hours the distribution of livor is not altered by changing the position of the body

The blood clots shortly after circulation ceases and when the process is slow the various elements of the blood settle out in a layered fashion The *dependent portion is a deep red while the yellow* upper portion resembles chicken fat and is called a chicken fat clot

Algor mortis The rate at which algor mortis, the cooling of the body occurs is dependent on many environmental conditions including temperature, protection from wind sun, and rain, the amount of clothing worn, and the presence of infection in the body. Under average conditions the body cools at a rate of 3–3 5°F for the first few hours the rate gradually decreasing until the environmental temperature is reached

Autolysis Autolysis is the breakdown of tissue by enzymes liberated by that tissue after death The rate of autolysis varies greatly with the tissue involved, being extremely rapid in the pancreas and the lining of the gallbladder and digestive tracts In autolysis, the nuclei of cells lose their staining ability, the cytoplasm assumes a homogeneous acidophilic granular appearance and with the further passage of time, cellular outlines may be lost *See* ENZYME

Putrefaction Finally putrefaction, the wide spread invasion of the body by organisms from the gastrointestinal tract, ensues Gas produced by the bacteria gives the tissues crepitance and an offensive odor There is often a greenish discoloration of the organs owing to staining by blood pigments which have been altered by the bacterial action

Necrobiosis. This is the physiological death of cells which normally occurs throughout the life history of the organism Examples may be found in the skin where there is constant maturation death, and replacement of cells and in the blood stream where the cells, after a relatively brief life span are destroyed and replaced by new ones See HEMATOPOIESIS

Necrosis. The pathological death of cells or tissues in a living organism is called necrosis Cells that are dead have lost the capability for respiration metabolism and reproduction Necrosis is classified by the abnormal gross or microscopic appearance of the cells or tissues involved this appearance is due to the action of enzymes mainly proteases Cells that have become necrotic fail to stain properly The nucleus may lose its affinity for the basic dyes used in routine histology so that on microscopic observation it is weakly visible if at all Alternatively, the nucleus may shrink and stain with greater intensity than usual (pyknosis) or it may even fragment Cells that have been dead for a considerable period of time assume a homogeneous acidophilic appearance Subsequently the cell membranes disintegrate so that no architecture is discernible Finally, the tissue undergoes autolysis and is resorbed or phagocytized

Coagulation necrosis In coagulation necrosis the general over all microscopic architecture is preserved but masses of fibrin or fibrinoid material are deposited in the necrotic mass so that the general microscopic appearance is that of a coagulative process This type of necrosis is commonly seen following obstruction of the blood supply to a tissue

Caseous necrosis Caseous or "cheesy" necrosis is so named because of the gross resemblance to cheese Usually no architecture can be discerned microscopically, although it is not uncommon to find the intercellular reticulum persisting for surprising periods of time after cell death in specially stained preparations Active tuberculosis and certain fungus infections are characterized by foci of caseous necrosis In these cases the necrosis is attributed to the action of bacterial or fungal toxins The spirochete of syphilis produces an infection which is often characterized by the presence of gummatous necrosis, a type of necrosis closely resembling caseous necrosis but differentiated by a more rubbery quality

Liquefaction necrosis In liquefaction necrosis, the necrotic mass first softens and subsequently liquefies Tissue death in the central nervous system usually results in this type of necrosis Shortly after an infarct of the brain occurs, there is cerebral softening Later the dead tissue liquefies so that eventually only a cystic defect remains

Fat necrosis Fat necrosis occurs in superficial tissues as the result of trauma circulatory disturbances or the injection of toxic substances It is also encountered in the abdominal cavity as the result of the release of pancreatic enzymes In fat necrosis the fat cells die and release fat into the intercellular space where it is changed into fatty acids and calcium soaps The presence of these insoluble calcium soaps gives the lesions their white to white yellow nodular appearance

Occurrence of necrosis. Necrosis is most commonly encountered clinically as the result of the obstruction of blood supply (ischemic necrosis) or infection Ischemic necrosis commonly manifests itself in three guises infarct decubitus, and gangrene In an infarct the necrotic portion is completely surrounded by living tissue The region of infarct is defined by the distribution of blood supply from the obstructed vessel Usually, infarcts are conical in shape with the apex of the cone marking the point of vascular obstruction At first an infarct appears red and congested Later it becomes yellow or yellow tan and firm bulging above the surrounding living tissue At this stage the microscopic appearance is that of coagulative necrosis With the passage of time the lesion becomes white and depressed as the necrotic cells are replaced by scar tissue

A decubitus or decubitus ulcer is necrosis of the skin and underlying tissues over a bony prominence in an individual who has lain in one position for such a long period of time that the circulation to these tissues is impaired by pressure between the bed and the bone Usually there is ulceration of the skin and the ensuing infection may contribute greatly to further extension of the necrotic area

In gangrene the necrotic tissue is not completely surrounded by living tissue and therefore is found most commonly in the nose the appendix or one of the extremities Gangrenous tissue gradually becomes yellow green then dark brown or black and is sharply demarcated from the adjacent healthy

grene which results from massive bacterial infection and necrosis Moist gangrene may occur in any part of the body Gas gangrene is a variety of moist gangrene in which one of the bacterial invaders (usually *Clostridium welchii*) produces gas and the involved tissue becomes honeycombed with bubbles See GANGRENE, GAS

Physical agents capable of causing necrosis include injury, heat freezing, electricity, ultraviolet light, x rays and other forms of ionizing radiation The list of chemical poisons which can produce necrosis is virtually endless Certain chemicals, such as chloroform have a selective effect of certain organs, others, such as the corrosive acids and alkalies, are effective in producing necrosis on contact with any tissue [W F P]

Decapoda (Crustacea)

One of the more highly specialized orders of the class Crustacea This order includes the shrimps, lobsters hermit crabs and true crabs The order is so diverse that satisfactory definition is difficult, but a few characters are common to nearly all decapods The most obvious is a head shield, or carapace which covers and coalesces with all of the thoracic somites and which overhangs the gills on each side The first three of the eight pairs of thoracic appendages are specialized as maxillipeds and closely associated with the true mouth parts The gills are usually well developed and arranged in several series

Decapods vary in size from less than one half inch to the great length of the giant Japanese crab, the largest living arthropod, a spider crab which may span more than 12 ft between the tips of the outstretched claws Although most decapods are found in the sea, they are by no means restricted to that habitat Crayfishes are well known inhabitants of fresh water streams and ponds, as are several kinds of shrimps and some true crabs A number of crabs and hermit crabs have become well adapted to a terrestrial existence far from water, they return to the sea only seasonally to hatch their eggs Many crayfishes burrow in the ground, and one species of crab spends its entire life in the tops of lofty trees

In a general way, the decapods may be divided into two groups the long tailed and the short tailed forms The long tailed species, such as the shrimps and lobsters have a more or less cylindrical or laterally compressed, carapace that often bears a head spine or rostrum The large, muscular abdomen permits the shrimp, or lobster, to dart quickly backward out of danger but, conversely, the succulence of this tail makes the animals prey to numerous enemies including man In the short tailed species the crabs, the carapace is often broadened and flattened dorsally and usually does not form a rostrum The much reduced and feeble abdomen is tucked under the thorax where it serves the female as a brood pouch for the eggs

Appendages. Like other malacostracans, nearly all decapods have 20 pairs of appendages (referred to by number in brackets) The eyestalks [1] are composed of 2 rarely 3, segments, either of which occasionally, may be abnormally

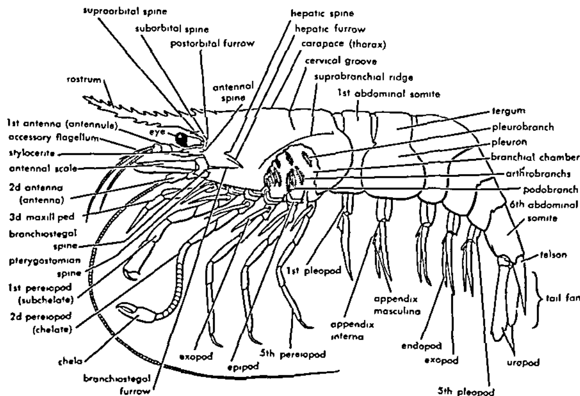


Fig 1 Lateral view of caridean prawn showing external morphology

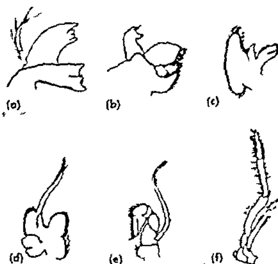


Fig 2 (a) Mandible of caridean prawn, Palaemon, showing incisor and molar processes and palp (b) First maxilla (maxillule) of same (c) Second maxilla (maxilla) of same (d) First maxilliped of same (e) Second maxilliped of same (f) Third maxilliped of same (The Smithsonian Institution)

elongate The eyestalks of the decapods therefore, are more like typical arthropod limbs than those of most other crustaceans The first pair of antennae the antennules [2] has a 3-segmented peduncle and, typically 2 flagella, the outer one being bifurcated near the base in some shrimps The second antennae [3] have up to 5 segments in the peduncle, a single flagellum and often a scalelike outer branch or exopod (Fig 1)

There are 6 pairs of mouthparts in the decapods, the posterior 3 pairs being modified thoracic legs The mandibles [4] sometimes have distinct incisor and molar processes These are sometimes indistinguishably fused, or the incisor process may be lacking There is usually a 3-segmented palp, but never a distinct lacinia mobilis as in most of the other malacostracans (Fig. 2) The first maxillae (maxillulae) [5] second maxillae (maxillae) [6], and first maxillipeds [7] are usually foliaceous The second maxillipeds [8] are also composed of broad, flattened segments but they are less modified from the typical thoracic appendages than the first maxillipeds The third maxillipeds [9] are elongate and leglike in the shrimps but are greatly modified forming an operculum over the preceding mouthparts in most of the crabs

The 5 remaining pairs of thoracic appendages, the pereiopods [10–14], are basically walking legs although 1 or more pairs may be variously modified The most common modification is seen in the formation of pinching claws, or chelae In the long tailed decapods 1–3, rarely 4 or all 5 pairs of pereiopods may be chelate In the short tailed forms only the first pair is usually thus modified Some of the pereiopods may be broadened and flattened as swimming paddles especially in the swimming crabs In the shrimps, the pereiopods are basically composed of 7 segments, of which some are occasionally multiarticulate In most of the other decapods on the other hand, the second and third segments are fused Occasionally, 1 or 2 pairs of pereiopods are vestigial or entirely lacking.

The first 5 pairs of abdominal appendages, the pleopods [15–19] are swimming organs in most of the long tailed decapods but the first 1–2 pairs

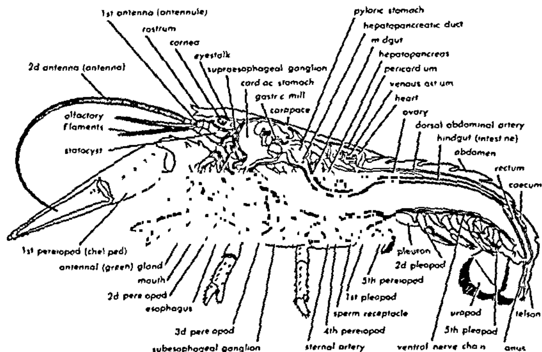


Fig 3 Median section of lobster, Homarus, showing general anatomy

grene which results from massive bacterial infection and necrosis. Moist gangrene may occur in any part of the body. Gas gangrene is a variety of moist gangrene in which one of the bacterial invaders (usually *Clostridium welchii*) produces gas and the involved tissue becomes honeycombed with bubbles. *See* GANGRENE GAS.

Physical agents capable of causing necrosis include injury, heat, freezing, electricity, ultraviolet light, x rays, and other forms of ionizing radiation. The list of chemical poisons which can produce necrosis is virtually endless. Certain chemicals, such as chloroform, have a selective effect of certain organs; others, such as the corrosive acids and alkalies, are effective in producing necrosis on contact with any tissue. [W.F.P.]

Decapoda (Crustacea)

One of the more highly specialized orders of the class Crustacea. This order includes the shrimps, lobsters, hermit crabs, and true crabs. The order is so diverse that satisfactory definition is difficult, but a few characters are common to nearly all decapods. The most obvious is a head shield, or carapace, which covers and coalesces with all of the thoracic somites and which overhangs the gills on each side. The first three of the eight pairs of thoracic appendages are specialized as maxillipeds and closely associated with the true mouth parts. The gills are usually well developed and arranged in several series.

Decapods vary in size from less than one-half inch to the great length of the giant Japanese crab, the largest living arthropod, a spider crab which may span more than 12 ft between the tips of the outstretched claws. Although most decapods are found in the sea, they are by no means restricted to that habitat. Crayfishes are well known inhabitants of fresh-water streams and ponds, as are several kinds of shrimps and some true crabs. A number of crabs and hermit crabs have become well adapted to a terrestrial existence far from water; they return to the sea only seasonally to hatch their eggs. Many crayfishes burrow in the ground, and one species of crab spends its entire life in the tops of lofty trees.

In a general way, the decapods may be divided into two groups: the long-tailed and the short-tailed forms. The long-tailed species, such as the shrimps and lobsters, have a more or less cylindrical or laterally compressed, carapace that often bears a head spine or rostrum. The large, muscular abdomen permits the shrimp, or lobster, to dart quickly backward out of danger but, conversely, the succulence of this tail makes the animals prey to numerous enemies including man. In the short-tailed species, the crabs, the carapace is often broadened and flattened dorsally and usually does not form a rostrum. The much reduced and feeble abdomen is tucked under the thorax where it serves the female as a brood pouch for the eggs.

Appendages. Like other malacostracans, nearly all decapods have 20 pairs of appendages (referred to by number in brackets). The eyestalks [1] are composed of 2, rarely 3, segments, either of which, occasionally, may be abnormally

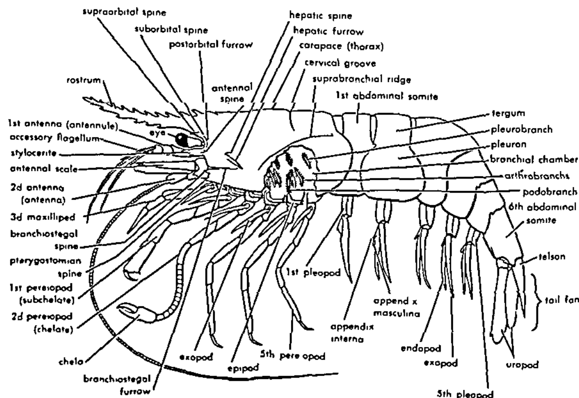


Fig. 1. Lateral view of caridean prawn showing external morphology.

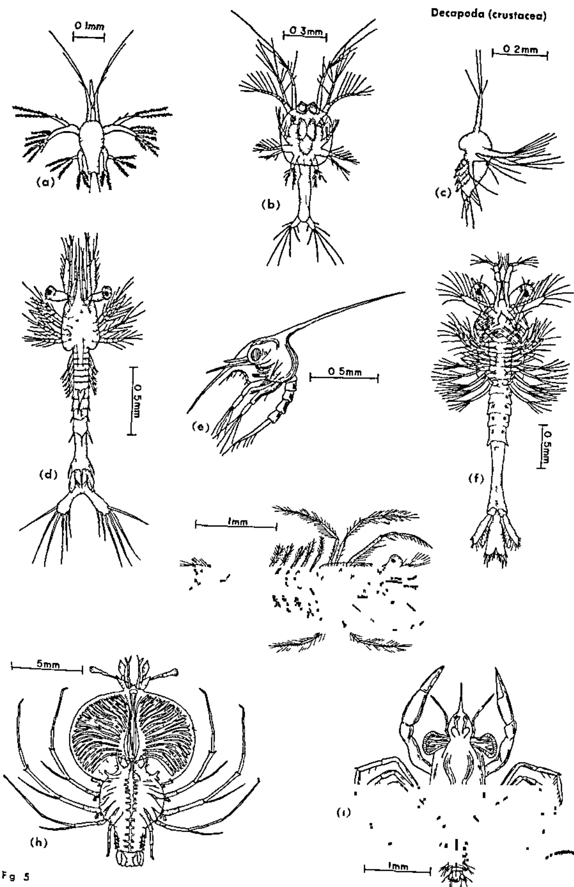


Fig 5

membrane of the branchial chamber This chamber is covered with minute villi is unusually well supplied with blood vessels and thus functions as a lung A similar highly vascularized area occurs in the delicate abdominal skin of the terrestrial hermit crabs

Excretion The antennal gland is the chief excretory organ in all decapods Maxillary glands are often present in the larval stages but they never persist in the adult The antennal gland is compact and more complex than in any other crustacean group It consists of three divisions the saccule which is usually partitioned or ramose the labyrinth a spongy mass with a complex system of canals and the bladder with a duct leading to the exterior The bladder may be a simple vesicle with a short duct as in the crayfish or it may be variously lobate as in most of the shrimps crabs and hermit crabs In the latter case two diverticula from the bladder extend through the entire abdomen The external opening is usually in a small elevation on the first segment of the second antennae In the crabs the aperture is covered by an operculum which may be opened and closed

Reproductive system The sexes are separate in most decapods although protandric hermaphroditism occurs in some shrimps The testes most often lie partly in the thorax and partly in the abdomen and they are usually connected across the midline In some hermit crabs however they lie wholly within the abdomen and they may be either unconnected or fused into a single organ The vas deferens usually opens on the first segment of the last pair of thoracic legs but in some crabs it perforates the last thoracic sternum The ovaries resemble the testes in shape and position The oviducts usually open on the first segment of the third pereiopods in most of the crabs however they open on the sternum

Development In only 1 group of shrimps are all of the presumably primitive larval stages represented (Fig 5) The first stage is a typical crustacean nauplius with an unprotected oval body bearing a median eye and 3 pairs of appendages This stage is followed by a metanauplius in which 4 more appendage rudiments appear The third pair of the original appendages becomes less of a swimming organ and more like a pair of primitive mandibles In the third stage the protozoea the 7 pairs of appendages become more highly developed a carapace covers the anterior part of the body the abdomen is clearly formed though unsegmented at first the rudiments of paired eyes appear and the heart is formed In the next stage the zoea the eyes become movable and the carapace develops a rostrum All thoracic appendages are present at least in rudimentary form and those of the abdomen appear especially the uropods In the last larval stage the mysis the well developed thoracic appendages replace the antennae as the chief swimming organs The abdomen increases markedly in size and takes on a form similar to that of the adult

In most decapods the naupliar stages are completed while in the egg and the animals hatch as zoeae In many the mysis stage is replaced by larvae or postlarvae of different kinds In the crabs the last zoea or metazoea metamorphoses into a megalopa This is similar to the adult except that the abdomen is prominent and the pleopods are used for swimming

The larval stages are even further reduced and sometimes completely eliminated in the fresh water decapods Most fresh water shrimps hatch in either a late zoeal or a mysis stage, and the crayfishes and fresh water crabs hatch as miniature adults The terrestrial forms on the other hand hatch in the sea eggs which then pass through the usual larval stages of marine species

Phylogeny No true decapod fossils are known with certainty from Paleozoic deposits but decapod shrimps are not uncommon in Triassic and especially Jurassic rocks Lobsterlike forms related to Recent deep sea species also appeared in the Triassic true lobsters are known from the Jurassic and crayfishes from the Cretaceous Crabs are well known as fossils with primitive groups appearing in the Jurassic, several more specialized groups are found in the Cretaceous and all of the major existing groups were represented in the Tertiary

The evolution of the decapods has not been satisfactorily worked out and may never be known completely The presence of a nauplius larva in the development of some shrimps suggests a link with the primitive crustaceans but it can hardly be doubted that the Decapoda as a group are highly specialized They are certainly closely related to the Euphausiacea so closely in fact that fusion of the two groups has been suggested by some authors

Classification The classification of the 8500 living species of decapods presents a difficult problem that has not yet been entirely solved The division of the order into long tailed and short tailed groups alluded to previously was in favor in the eighteenth century This scheme was generally abandoned however when it was discovered that some of the long tailed forms were obviously more closely related to the short tailed group than they were to other long tailed species The chief disparity seems to lie between the shrimps (Natantia) and all of the other decapods (Reptantia) as follows

Order Decapoda
- Suborder Natantia
  - Section Penaeidea
  - Section Caridea
  - Section Stenopodidea
- Suborder Reptantia
  - Section Macrura
  - Section Anomura
  - Section Brachyura

Penaeidea The primitive section of the Decapoda penaeideans (Fig 6) are distinguished by

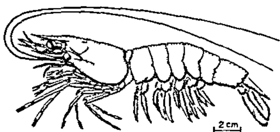


Fig 6 Penaedean prawn Penaeus setiferus (The Smithsonian Institution)

the following characters the pleura of the second abdominal somite overlap those of the first the third legs are nearly always chelate but are no stouter than the first pair the first pleopods of the male bear a complicated flaplike appendix the petasma the females have a spermatophore receptacle the thelycum on the ventral surface of the posterior thoracic somites and the gills are dendrobranchiate Unlike all other decapods the eggs are not attached to the abdominal appendages but are almost always shed free into the water The Penaeidea share with the Euphausiacea the presence of a petasma in both groups the eggs are rarely carried and they hatch at an early larval stage usually as nauplii The best known penaeideans including most of the commercially important shrimps or prawns of the warmer seas (Penaeidae) live on muddy bottoms in shallow or moderate depths However they also occur both in the deep sea and as pelagic organisms in the mid depths In this latter region are found the Sergestidae in which the last two pairs of thoracic legs are reduced or absent There are more than 300 species of living penaeideans

Caridea This is the largest and most diverse group of shrimps and prawns In this group the pleura of the second abdominal somite overlap those of the first the third legs are never chelate there is no petasma or thelycum and the gills are phyllobranchiate The sexes can usually be distinguished by the presence in the male of two stylets an appendix masculina and an appendix interna located on the inner edge of the inner branch of the second pleopods only the appendix interna occurs in the female Carideans are found in all parts of the sea often in association with other marine animals Members of at least two families the Atyidae and Palaemonidae are also widespread in fresh water The edible shrimps and prawns of northern Europe and of northwestern North America of the genera *Crangon* and *Pandalus* belong to the Caridea as do the large fresh water prawns of the tropical genus *Macrobrachium* Some of the latter are the largest natantians known attaining a length of more than a foot and having abnormally long second pereiopods which may be fully as long as the body (Fig 7) The snapping shrimps or Alpheidae make a loud popping noise with a socket and plunger mechanism on the overdeveloped first pereiopods sometimes interfering seriously with underwater sound equipment in tropic seas Most shrimps and prawns are favorite foods of fishes but some carideans maintain "service stations" for fishes where they remove parasites from the outer skin mouths and gills of their customers More than 1500 species of Caridea are known

Stenopodidea This is a small group of shrimps which superficially resemble the Penaeidea (Fig 8) The third pereiopods are chelate but are much longer and stouter than the first pair the pleura of the second abdominal somite do not overlap those of the first there is no petasma or thelycum and the gills are trichobranchiate All of the 20 or more living species are marine some are closely associated with other animals such as sponges

Macrura This long tailed group of the Reptantia includes the deep water and fossil eryonids the spiny lobsters the true lobsters and the mud

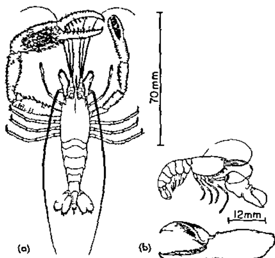


Fig 7 (a) River prawn Macrobrachium faustinum (b) Snapping shrimp Alpheus heterochaelis with cheliped enlarged (The Smithsonian Institution)

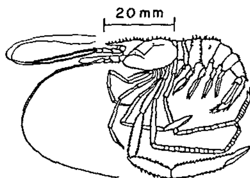


Fig 8 Stenopodid shrimp Stenopus hispidus (Smithsonian Institution)

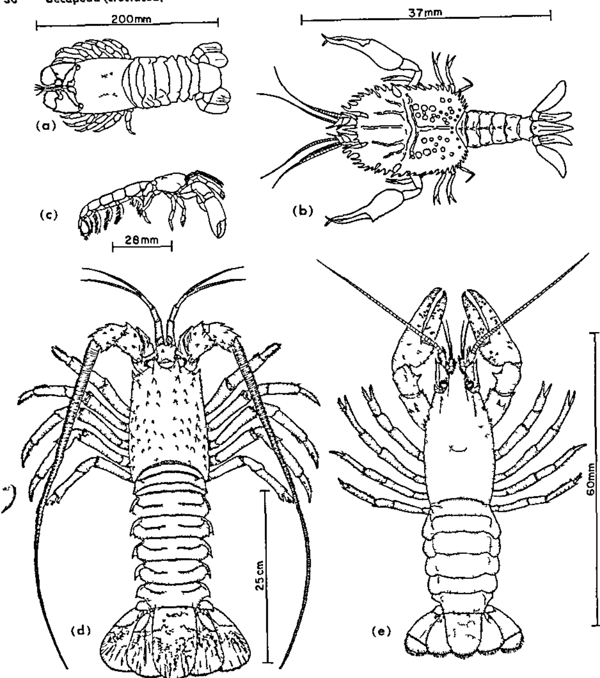


Fig 9 (a) Span sh lobster *Scyllar des aequ noct al s* (b) Eryonid *Polycheles cruc fer* (c) Mud shrimp *Call an dea laev cauda* (d) Sp ny lobster *Panul rus nterruptus* (e) Crayf sh *Orconectes l mosus* (The Sm thson an Inst tut on)

shrimps The abdomen is extended and bears a well developed tail fan There are about 700 known living species (Fig 9)

The Fryonidea are rather thin shelled blind in hal itants of the depths of the sea The carapace is flattened dorsally and considerably expanded lat erally The first four or all five pairs of pereio-pods are chelate with the first pair much elon gated There about 40 living species of Eryonida, the group was probably more numerous in ancient seas

The Scyllaridea include the spiny lobsters or langoustes (Palinuridae) and the Spanish or shovel nosed lobsters (Scyllaridae) They are heavily armored like the true lobsters but are dis tinguished by the absence of a rostrum and of che-lae except occasionally on the last pereiopod of the female They are abundant in shallow and mod erate depths of warm and temperate seas where they are often of considerable commercial impor tance Frozen tails of spiny lobsters are imported to America from Cuba South Africa Australia

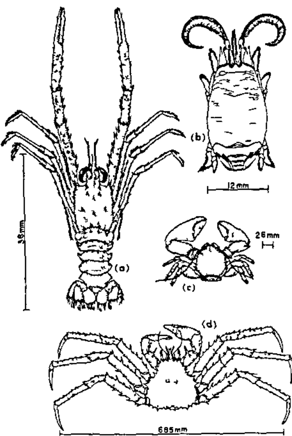


Fig 10 (a) Galatheid Munida evermanni (b) Mole crab Emerita talpoida (c) Porcellanid crab Petrolisthes tridentatus (d) King crab Lithodes maia (The Smithsonian Institution)

and New Zealand and the fresh meat supports fisheries in southern Florida southern Europe and Japan There are about 85 Recent species

The true lobsters and crayfishes Nephropsidea also have a firm shell but are also characterized by a rostrum and by chelae on the first three pairs of pereiopods with the first pair being noticeably larger than the others Most lobsters (Nephropsidae) are found in cool seas or in the cool offshore waters of the tropics The most familiar lobsters (*Homarus*) are those found along the Atlantic coasts of Europe and North America The Norway lobster of Europe (*Nephrops*) is also of some commercial importance but is smaller and less meaty The crayfishes (Cambaridae Astacidae Parastacidae) are widespread in fresh waters of the temperate regions of all continents except Africa They are of commercial importance in southern Europe and Australia More than 300 living species of lobsters and crayfishes are known more than half of them being from the fresh waters of the United States

The mud shrimps Thalassinidea are usually thin shelled burrowing crustaceans with large chelate or subchelate first pereiopods and no chelae on the third pereiopods More than 250 living species are found in shallow and deep seas throughout the world especially in tropical and warm temperate regions

Anomura This intermediate and possibly unnatural group lies between the Macrura and the Brachyura It includes the galatheids the porcelain crabs or rock sliders the hermit crabs the king crabs and the mole crabs or hippas In nearly all of these diverse crustaceans the first and sometimes the last pereiopods are chelate or subchelate The abdomen is usually bent forward ventrally or is asymmetrical soft and twisted There are about 1300 Recent species

The Galatheidea (Fig 10) includes those anomurans with a symmetrical abdomen which is bent upon itself and provided with a well developed tail fan Most of the approximately 575 living species are found in the sea at shallow or considerable depths but one aberrant genus (*Aegla*) inhabits fresh water streams of temperate South America The more primitive species (Galatheidae) are somewhat lobsterlike the rock sliders (Porcellanidae) resemble the true crabs but may be distinguished from them by the much reduced and chelate fifth pereiopods and by the well developed tail fan They are most abundant in the intertidal zone and in moderate depths of the warmer seas

The Paguridea include the hermit and king crabs in which the abdomen is nearly always asymmetrical being either soft and twisted or bent under the thorax The uropods when present as in the hermit crabs are adapted for holding the body in a portable shelter The first and often the last pereiopods are chelate The hermit crabs (Fig 11) Paguridae are found everywhere in the sea and are especially abundant in the intertidal zone A few species (Coenobitidae) are also found on land in the tropics Most hermit crabs live in the shells of dead gastropod mollusks which

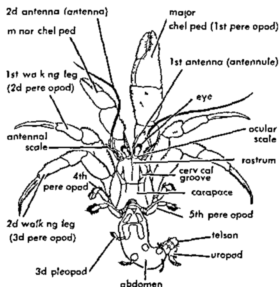


Fig 11 Dorsal view of hermit crab *Pagurus* showing external morphology

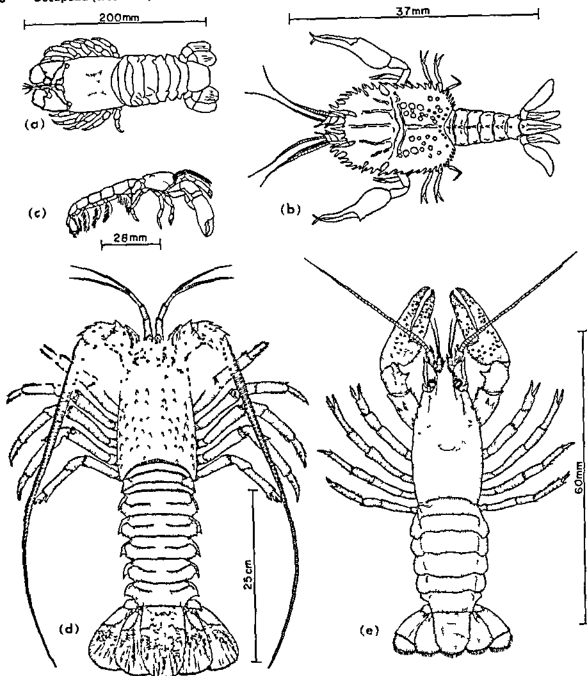


Fig 9 (a) Spanish lobster *Scyllarides aequinoctialis* (b) Eryonid *Polycheles crucifer* (c) Mud shrimp *Callianidea laevicauda* (d) Spiny lobster *Panulirus interruptus* (e) Crayfish *Orconectes limosus* (The Smithsonian Institution)

shrimps The abdomen is extended and bears a well developed tail fan There are about 700 known living species (Fig 9)

The Eryonidea are rather thin shelled blind inhabitants of the depths of the sea The carapace is flattened dorsally and considerably expanded laterally The first four, or all five pairs of pereiopods are chelate, with the first pair much elongated There about 40 living species of Eryonida, the group was probably more numerous in ancient seas

The Scyllaridea include the spiny lobsters or langoustes (Palinuridae) and the Spanish, or shovel nosed lobsters (Scyllaridae) They are heavily armored like the true lobsters but are distinguished by the absence of a rostrum and of chelae, except occasionally on the last pereiopod of the female They are abundant in shallow and moderate depths of warm and temperate seas where they are often of considerable commercial importance Frozen tails of spiny lobsters are imported to America from Cuba, South Africa, Australia

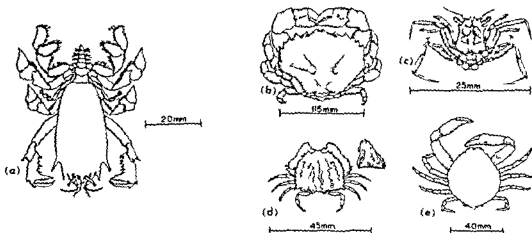


Fig 13 (a) Gymnopleuran crab, *Raninoides louisianensis* (b) Dromiid crab *Dromia erythropus* (c) Mask crab *Ethusa mascarone americana* (d) Box crab *Calappa sulcata* (e) Purse crab *Persephona punctata* (The Smithsonian Institution)

is not adapted for swimming, and the first pereiopods are either styliform or subchelate The common hippas live in sand in the surf zone of tropical and temperate shores where they move up and down the beach with the tide, they are often used as bait

Brachyura These are the true crabs (Fig 12) The abdomen is symmetrical without a tail fan and bent under the thorax The first pereiopods are always chelate or subchelate The true crabs are as numerous as all other decapods combined numbering nearly 4500 Recent species They are divided into four subsections

The Gymnopleura include about 30 species of primitive burrowing crabs, their carapaces are more or less trapezoidal or elongate the first pere-

Fig 14 (a) Spider crab *Mithrax acuticornis* (b) Fiddler crab *Uca pugilator* (c) Land crab *Gecarcinus lateralis* (d) Fresh-water crab *Epilobocera sinuatifrons* (e) Mud crab *Eurypanopeus abbreviatus* (f) Swimming crab *Ovalipes ocellatus* (g) Stone crab, *Menippe mercenaria* (The Smithsonian Institution)

iopods are subchelate and some or all of the remaining pereiopods are flattened and expanded for burrowing (Fig 13)

The Dromiacea is another primitive group of about 200 species The first pereiopods are chelate the last pair are dorsal in position and modified for holding objects such as sponges tunicates, and bivalve mollusk shells over the crab The oviducts open on the first segments of the third pereiopods rather than on the sternum as in most brachyurans The mouth frame is quadrate

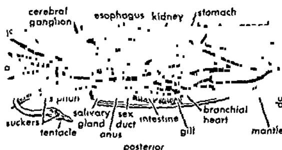
In Oxystomata the first pair of pereiopods are chelate and the last pair are either normal or modified as in the Dromiacea The mouth frame is triangular and forward over the epistome The oxystomes include the mask crabs (Dorippidae) the box crabs (Calappidae) and the purse crabs (Leucosiidae) There are nearly 500 Recent species

Most of the crabs more than 3700 living species belong to the Brachygnatha (Fig 14) in which the mouth frame is quadrate and the last pereiopods are rarely reduced or dorsal in position To this group belong the swimming crabs (Portunidae) in which the last pereiopods are modified as swimming paddles A well known swimming crab is the edible blue crab (*Callinectes*) of the shallow waters of the tropical and temperate Americas Also worthy of mention are the fresh water crabs (Potamonidae) which are found in tropical and some temperate regions of the world usually in those areas not inhabited by crayfishes The crabs of the genus *Cancer* are large and abundant enough to be of commercial importance in northern Europe and North America The rock or Jonah crabs of New England and the Dungeness crab of the Pacific coast are familiar edible crabs of this genus The ubiquitous and perplexing little mud crabs (Xanthidae) are well known to every visitor to rocky shores One of them the stone crab (*Menippe*) is highly esteemed by connoisseurs of sea food The pea crabs (Pinnotheridae) are often found in oysters and other bivalve mollusks The square backed crabs (Grapsidae) are characteristic of warm marshy areas but are not restricted to that habitat they mark the trend toward the true land crabs (Gecarcinidae) the depredations and intrusions of which are familiar to all who live in the tropics The ghost crabs (*Ocypode*) which scuttle almost unseen over sandy beaches and the fiddler crabs (*Uca*) of muddy shores are closely related The end of the list is reached with the spider or decorator crabs (Majidae) slow moving animals that often conceal themselves by attaching seaweeds and various sessile animals to their carapaces One of them (*Macrocheira*) is the largest arthropod now alive *See* CRUSTACEA [F A C]

Bibliography H Balss *Decapoda*, in W Kukenthal and T Krumbach *Handbuch der Zoologie*, vol 7, pt 1, 1927, H Balss E Korschelt and W, v Buddenbrock, *Decapoda* in H G Bronn (ed), *Klassen und Ordnungen des Tierreichs*, vol 5, pt 1, 1940–1957, W T Calman, *Crustacea*, in R Lankester (ed), *A Treatise on Zoology*, pt 7, fasc 3, 1909

Decapoda (Mollusca)

An order of dibranchiate cephalopods, containing the squids and cuttlefishes They are characterized by the possession of eight arms and two long, often retractile, tentacles The cuttlefishes, *Sepia*, are confined to the Old World Their shells are used in the manufacture of dentifrices and cosmetics and artist's sepia is made from the ink The true squids are slender, swift swimmers, and some are of economic importance Two natural groups occur The Myopsida, represented by *Loligo* (see illustration), have the eye covered by the skin of the head



The squid, Loligo internal structure as seen with body wall and arms removed on left side (From T I Storer and R L Usinger, General Zoology, 3d ed , McGraw Hill 1956)

and are mainly coastal species The oceanic Oegopsida have the eyeball exposed to the water and the suckers may be modified into hooks Both the giant *Architeuthis* and *Illex* belong to this group *See* CEPHALOPODA DIBRANCHIA [G L V]

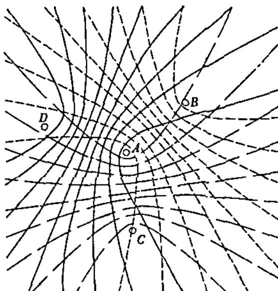
Bibliography G L Voss A review of the cephalopods of the Gulf of Mexico, *Bull Marine Sci Gulf and Carib* 6 85–178 1956

Decca

A hyperbolic navigation system which establishes a line of position from measurement of the phase difference between two continuous wave signals The intersection of the two lines of position from two pairs of transmitting stations establishes a navigational fix, or location Charts surprinted with Decca hyperbolas of constant phase difference are used to show the navigator's position Moving charts upon which a pen traces the track of the craft can be used to show position and provide a permanent record

The two lines of position required for a fix can be provided by one "master" and two "slave" transmitters A third slave located at the apex of a triangle in which the master is centered is also used

Since continuous waves are employed each station must use a different radio frequency Each



Decca station location and patterns A Master station B C D slave stations (From P C Sandretto Electronic Aviation Engineering International Telephone and Telegraph Corp 1958)

slave frequency is chosen to have a fixed ratio to that of the master transmitter If a fundamental (nontransmitted) frequency for a chain is approximately 14 17 kilocycles (kc) the table shows the radiated frequencies and their harmonic relationships for a typical chain

Master and slave station frequency relationship

Station	Frequency kc	Harmonic
Master	85 000	6
Red slave	113 333	8
Green slave	127 500	9
Purple slave	70 833	5

Signals from the master transmitter are picked up by each slave station and the phase of a multiple of the common harmonic is compared with that of the local slave transmission The slave transmission is locked in phase with that of the master

A receiver moved along the base line joining two Decca transmitters will show a continually changing phase relationship passing through zero many times between waves received from the slave station and those from the master transmitter Since no instrument can distinguish between 0° and 360° 30° and 390° and so on an adding type phase meter or decometer is used This meter rotates continuously and adds up the total number of degrees of phase shift The regions between lines of zero phase difference are known as lanes These lanes exist throughout the area served by the stations and vary in width diverging with distance from the base line

Since the phase relationship continually changes and passes through multiple zero positions means must be found to set the decometers at the proper reading in the event that receiver power is turned off or radio frequency signals are lost while the craft proceeds more than half a lane Although conventional astral or direction finder fixes can be employed to reset the decometers the Decca system provides for lane identification once a minute by substituting special signals that actuate an additional phase meter to show the lane within a so called zone in which the receiver is located The zones shown on navigational charts are all of equal width and each one comprises either 24 red 18 green or 30 purple lanes It has been considered unlikely that a craft would require identification of zones

Operating at radio frequencies in the order of 50–150 kc the useful range should be about 300 miles day or night Daytime distance may exceed 1000 miles While a theoretical accuracy of 1° of phase is possible it is doubtful that such accuracy can be attained in practice Published reports indicate daytime accuracy of 100 yd and nighttime accuracy of 500 yd at 300 miles *See* HYPERBOLIC NAVIGATION SYSTEM [A A M]

Decibel

A logarithmic unit used to express the magnitude of a change in level of power voltage current or sound intensity A decibel (db) is 1/10 bel (*see* BEL)

In acoustics a step of 1 bel is too large for most uses It is therefore the practice to express sound intensity in decibels The level of a sound of intensity I in decibels relative to a reference intensity I_R is

$$10 \log_{10} \frac{I}{I_R}$$

Because sound intensity is proportional to the square of sound pressure P the level in decibels is

$$10 \log_{10} \frac{P^2}{{P_R}^2} = 20 \log_{10} \frac{P}{P_R}$$

The reference pressure is usually taken as 0 0002 dynes/cm² or 0 0002 microbar (*see* SOUND PRESSURE) (The pressure of the earth s atmosphere at sea level is approximately 1 bar) A sinusoidal alternation in pressure at a frequency of 1000 cycles per second is barely audible to the average person when it has a root mean square sound pressure of 0 0002 microbar By this definition such a tone has a sound pressure level of 0 db

The neper is similar to the decibel but is based upon natural (Napierian) logarithms One neper is equal to 8 686 db *See* NEPER *see also* VOLUME UNIT (VU) [K D K]

Deciduous plants

Plants that regularly lose their leaves at the end of each growing season Dropping of the leaves occurs at the inception of an unfavorable

characterized by either cold or drought or both Most woody plants of temperate climates have the deciduous habit and it may also occur in those of tropical regions having alternating wet and dry seasons Many deciduous trees and shrubs of regions with cold winters become evergreen when grown in a warm climate Conversely such trees as magnolias evergreen in warm areas become deciduous when grown in colder climates *See* LEAF (BOTANY) PLANT PHYSIOLOGY PLANT TAXONOMY [N A]

Decimal number system

The system used in ordinary arithmetic in which numbers are represented as linear combinations of powers of 10 The base 10 has as its origin the fact that man has a total of 10 digits on both hands Other than this the decimal number system has little to recommend it over other systems *See* NUMBER SYSTEMS

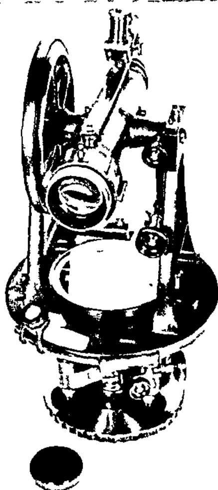
Declinometer

A geomagnetic instrument for measuring magnetic declination sometimes called variation of the compass which is defined as the angle between true or geographic north and magnetic north—the direction of the magnetic meridian The instrument is suited for determining the direction in which a perfect compass needle would point and for comparing this with the known true azimuth of a fixed reference point or a celestial object such as the sun or the North Star

Compass declinometer Used for many years by the U S Coast and Geodetic Survey for magnetic distribution surveys the compass declinometer employs a thin compass needle 6 in long supported on a sapphire bearing and steel pivot of high quality Peep sights serve for aligning the compass box on an azimuth mark The accuracy attained is 2–3 of arc in middle latitudes

Transit declinometer This is the usual instrument for declination field work A surveyor's transit built to exacting specifications with respect to freedom from traces of magnetic impurities and quality of the compass needle has a 17 power telescope for sighting on a mark and for making solar and stellar observations to determine *true directions* A *microscope mounted on the side* of the telescope permits a more accurate alignment of the instrument axis with the magnetic needle

Error corrections Both of these declinometers must be standardized by observing at a magnetic observatory where the declination is accurately measured with standard instruments before their results can be relied upon The principal sources of error are (1) excessive pivot friction the effect of which may become quite substantial in polar latitudes where the horizontal intensity of the earth's magnetic field is small and there is little directive force to influence the rest position of the compass needle (2) eccentric mounting of the pivot relative to the horizontal circle which may result in an error of several minutes of arc even



Transit declinometer (*U.S Coast and Geodetic Survey*)

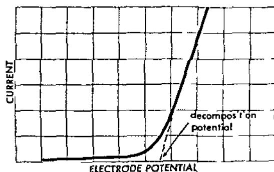
in an instrument of good quality and (3) lack of parallelism between the geometric and magnetic axes of the needle The accuracy of a properly calibrated transit declinometer in equatorial and middle latitudes is from 1 to 2 of arc

Precise observations For magnetic field work of highest precision and for standard calibrating instruments in a magnetic observatory a magnetometer is used to measure both declination and *horizontal* intensity This instrument employs a permanent magnet suspended by a fine gold ribbon and equipped either with a collimating lens and cross hair or with a mirror The position of the magnet is observed with a suitable telescope and scale which are integral parts of the equipment The suspension system eliminates pivot friction and the magnet itself may be inverted to offset the effect of nonparallelism of geometric and magnetic axes The telescope serves also for sighting on an azimuth mark Effects of torsion in the suspension fiber are eliminated by suspending a nonmagnetic weight to find the no torsion setting for the upper *fiber clamp Accuracy is 0'1–0'2 of arc for an* observatory declinometer [J H NF]

Bibliography *See* MAGNETOMETER

Decomposition potential

The electrode potential at which the electrolysis current begins to increase appreciably Decom position potentials are used as an approximate characteristic of industrial electrode processes *see* ELECTROCHEMICAL PROCESS, ELECTROLYSIS



Determinat on of decomposition potential

Decomposition potentials are obtained by extrap olation of current potential curves for a discus sion of these curves *see* OVERVOLTAGE Extrapola tion is not precise because there is a progressive increase of current as the electrode potential is varied The decomposition potential for a given element depends on the range of currents being considered The cell voltage at which electrolysis becomes appreciable is approximately equal to the algebraic sum of the decomposition potentials of the reactions at the two electrodes and the ohmic drop or voltage drop in the electrolytic cell The ohmic drop term is quite negligible for electro lytes with high conductance *See* ELECTROLYTIC CONDUCTANCE [P D]

Decompression illness

Symptoms in man which result from a sudden re duction in atmospheric pressure It is also called decompression sickness caisson disease the bends and compressed air illness It is most commonly seen in two groups of subjects (1) those who rapidly ascend in nonpressurized airplanes to alti tudes in excess of 18 000 ft and (2) among divers and sand hogs who work under increased atmos pheric pressure such as compressed air

The onset of symptoms may occur at any time from a few minutes to several hours after decom pression The most common manifestation is pain in the joints and muscles However skin respira tory and neurologic symptoms are not uncommon The skin manifestations are itching discoloration and edema (swelling) Respiratory symptoms are coughing and dyspnea (difficulty in breathing) The neurologic symptoms are of a more grave nature and vary from mild paresthesia (sensations of tingling crawling or burning of the skin) and weakness to total paralysis loss of bladder and rectal sphincter control is common The severe forms of this illness are followed by circulatory failure paralysis coma and death

This condition is caused by the formation of nitrogen bubbles in the tissues and blood vessels In the body the nitrogen normally dissolved in body fluids forms bubbles when the atmospheric pressure is reduced These bubbles plug vessels and expand in tissue spaces such as muscles and joints pro ducing the characteristic symptoms and signs of this illness Helium when used by divers can also produce a similar condition

Treatment is recompression followed by gradual decompression to normal atmospheric pressure Prognosis is generally good except in those sub jects which show central nervous system damage *See* AVIATION MEDICINE BIOPHYSICS SPACE BIOL OGY [M CO]

Decontamination (radioactive contaminants)

The removal of radioactive contamination which is deposited on surfaces or may have spread through out a work area Personnel decontamination is also included The presence of radioactive contamina tion is a potential health hazard and in addition it may interfere with the normal functioning of plant processes particularly in those plants using radia tion detection instruments for control purposes Thus the detection and removal of radioactive con taminants from unwanted locations to locations where they do not create a health hazard or inter fere with production are the basic purposes of de contamination

There are four ways in which radioactive con tamination adheres to surfaces and these limit the decontamination procedures which are applicable The contamination may be (1) lying loosely on the surface (2) absorbed in porous materials (3) ab sorbed on or by the surface in the form of ions atoms or molecules or (4) mechanically bonded to surfaces through oil grease tars paint and so on

Methods of decontamination Decontamination methods follow two broad avenues of attack me chanical and chemical Commonly used mechanical methods are vacuum cleaning sand blasting blast ing with other abrasives flame cleaning scraping and surface removal (for example removal of concrete floors with an air hammer) The principal chemical methods of decontamination are water washing steam cleaning and scrubbing with de tergents acids caustics and solvents

Another important method of handling contami nation is to store the contaminated object or tem porarily abandon the contaminated space This can be done when the use of the material or space is not necessary for a period of time and the half life of the contaminant is relatively short For ex ample tools contaminated with short lived fission products may be stored or a bu nated with such material may l barred from use until the r

cay has reduced the contamination to an acceptable level

Other methods involve covering the contamination by some method such as painting and disposing of part or all of the contaminated equipment or facility Considerations which determine the methods used for decontamination or removal of contamination include (1) the hazards involved in the decontamination procedure (2) the cost of removal of the contamination and (3) the permanency of removal of the contamination (for example painting over a surface contaminated with a long lived radioactive material only postpones ultimate disposal considerations)

Personnel decontamination Personnel decontamination methods differ from those used for materials primarily because of the possibilities of injury to the person being decontaminated Procedures used for normal personal cleanliness usually will remove radioactive contaminants from the skin and the method used will depend upon its form and associated dirt (grease oil soil and so on) Soap and water (sequestrants and detergents) normally remove more than 99% of the contaminants If it is necessary to remove the remainder chemical methods which remove the outer layers of skin upon which the contamination has been deposited can be used These chemicals—citric acid potassium permanganate and sodium bisulfite are examples—should be used with caution and preferably under medical supervision because of the increased risk of injury to the skin surface he use of coarse cleansing powders should be [illegible]led for skin decontamination because they may lead to scratches and abraded skin which can permit the radioactive material to enter the body Similarly the use of organic solvents should be avoided for skin decontamination because of the probability of penetration through the pores of the skin It is very difficult to remove radioactive material once it is fixed inside the body and the ensuing hazard depends very little on the method of entry into the body that is through wounds through pores of the skin by injection or by inhalation When certain of the more dangerous radioactive materials such as radium or plutonium have been taken into the body various chemical treatments have been attempted to increase the body elimination but the results of these treatments are not very encouraging In the case of plutonium and certain other heavy metals the most effective treatment for removal from the body is the administration of chelating agents such as calcium ethylenediaminetetraacetate (CaEDTA) or a sodium citrate solution of zirconyl chloride In any case the safest and most reliable procedure for preventing internal exposure from radioactive material is not the application of internal decontamination but rather the application of health physics procedures for the prevention of entry of radioactive material into the body

Air decontamination Air contaminants frequently are eliminated by dispersion into the atmosphere Certain meteorological conditions such as prevailing wind velocities wind direction, and inversion layers seriously limit the total amount of radioactive material that may be released safely to the environment Consequently decontamination of the air stream by filters cyclone separators scrubbing with caustic solutions and entrapment on charcoal beds is often resorted to The choice of method used is guided by such things as the volume of airflow the cost of heating and air conditioning the hazards associated with the airborne radioactive material and the isolation of the operation from other populated areas

Water decontamination Water decontamination processes can use one or both of the two opposing philosophies of maximum dilution or maximum concentration (and subsequent removal) of the contaminant Water concentration methods involve the use of water purification processes that is ion exchange chemical precipitation floculation filtration and biological retention

Certain phases of radioactive decontamination procedures are potentially hazardous to personnel Health physics decontamination practices include the use of protective clothing respiratory devices localized shielding isolation or restriction of an area provisions for the proper disposal of the attendant wastes and application of the recommended rules and procedures for limiting the internal and external doses of ionizing radiation *See* MONITORING (IONIZING RADIATION), RADIATION INJURY (BIOLOGY) RADIOACTIVE WASTE DISPOSAL, RADIOACTIVITY [K Z M]

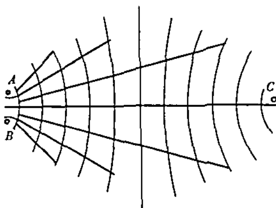
Bibliography *Control and Removal of Radioactive Contamination in Laboratories* Natl Bur Standards Handbook 48 S Kinsman *Radiological Health Handbook* U S Dept Health Education and Welfare PB 121784 1957

Dectra

A variation of the Decca principle designed to provide a track and distance measurement along a narrow track rather than a widespread navigational grid Its name is derived from Decca track and range"

A Dectra chain may comprise three stations located as in the diagram Stations A and B generate a family of hyperbolic position lines which by virtue of the relatively short base line (about 80 miles) are nearly straight over a major portion of their lengths Station A (the common master transmitter) and station C provide hyperbolic lines of position which at their intersections with the track pattern indicate a fix

Stations A and B transmit continuous waves on the same frequency of about 70 kilocycles The transmission from A is interrupted every few seconds for a fraction of a second During this recurrent silent period station B transmits a signal phase locked to that of A A line of position resulting from the phase difference measured at the receiver is indicated on conventional Decca equipment



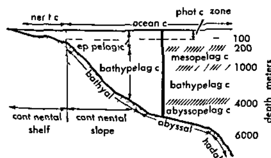
Three stat on Dectra cha n Track nformat on is fur n shed by A and B range s derived from A and C

The ranging pattern results from the pair A and C whose radio frequency transmissions are related to a common subharmonic frequency The mas ter and slave" transmissions are locked in phase *See* DECCA HYPERBOLIC NAVIGATION SYSTEM

[A A M]

Deep sea fauna

The deep sea may be regarded as that part of the ocean below the upper limit of the continental slopes (Fig 1) Its waters fill the deep ocean ba sins cover about two thirds of the earth s surface have an average depth of about 4000 meters (m) and provide living space for communities of ani mals that are quite different from those inhabiting the land fringing waters which overlie the continen tal shelves (neritic zone)



F g 1 Class f cat on of mar ne env ronments The r ght hand part of the d agram llustrates the recent proposal to d vide the bathypelag c zone into meso pelag c bathypelag c and abyssopelag c zones The d vis on of the deep sea benth c reg on nto bathyal abyssal and hadal zones also is shown

The systematic exploration of the deep sea be gan with the voyage of the HMS *Challenger* (1872 1876) Since that time there have been some 20 large scale deep sea expeditions Only recently, however have the deeper parts of the ocean below 6000 m been explored In 1948 the Swedish Deep Sea Expedition in the Atlantic developed new tech niques for trawling and the winch used later by the Danish *Galathea* Since 1949 the deep sea has been explored by Russian and in 1950–1952 by Danish research ships

The deep sea fauna consists of pelagic animals (swimming and floating forms between the surface and deep sea floor) and below these the benthos or bottom dwellers which live on or near the ocean bottom Pelagic animals can be divided into the usually smaller forms that tend to drift with the currents (zooplankton) and the larger and more ac tive nekton such as squids fishes and cetaceans Pelagic deep sea animals are frequently termed bathypelagic in contrast to the epipelagic or ganisms of the surface waters (Fig 1)

Bathypelagic fauna All animal life in the sea pelagic and benthic depends on the growth of mi croscopic plants (phytoplankton) From the sur face down to a maximum depth of about 100 m there is sufficient light for photosynthesis and vigor ous phytoplanktonic growth This layer is known as the photic zone In the deep sea plants can ex ist only as saprophytes The productivity of the plants however is reflected down to the deepest parts of the sea through complex food chains These consist of zooplankton that graze on phyto-plankton carnivorous species that feed on zoo plankton and large predators that eat the other animals The typical bathypelagic animals (Fig 2) as observed from bathysphere and bathyscaph begin to appear below depths of about 200 m

Zooplankton The planktonic or drifting forms of animal life in the ocean include the Protozoa lar val stages of deep sea fishes and even larger or ganisms with limited powers of movement

1 Protozoa Included in this group are various species of Foraminifera and Radiolaria such as Challengeridae and Tuscaroridae the skeletons of which form an important part of the deep sea sedi ments

2 Coelenterata Scyphomedusae such as *Atolla* and *Periphylla* are not uncommon Other jellyfishes include various Trachymedusae (*Crossota* and *Col obonema*) and Narcomedusae Siphonophora par ticularly the diphyids are found down to depths of at least 3000 m but are more common in the upper few hundred meters One family of Ctenophora (comb jellyfish) the Bathyctenidae is entirely bathypelagic

3 Nemertea This group of worms has bathype lagic species belonging to some 10 families

4 Crustacea In numbers of species and individ uals the small Copepoda (from 0 3 to 17 mm in length) are the dominant group of crustaceans in the ocean There are numerous bathypelagic spe-cies Certain of the Ostracoda (*Gigantocypris*) are purely bathypelagic as are some of the Amphipoda (the gammarid genera *Cyphocaris* and *Hyperiopsis* and most species of the hyperiid families Scinidae and Lanceolidae)

The larger and more active pelagic crustaceans (Euphausiacea various Mysidacea and prawns) are usually classed as plankton but might well be

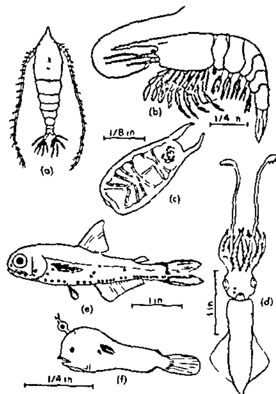


Fig 2 Pelagic animals of the deep sea (a) Copepod, *Haloptilus acutifrons* length 1–8 mm (b) Prawn *Acanthephyra multispina* (c) Salp, *Salpa (Thalia) democratica* (d) Squid *Abraliopsis morisii* (e) Lantern fish *yctophym punctatum* (f) Angler fish *Lophodolus anthognathus*

called "micronekton" a group intermediate between thrusting nekton and feebler swimming plankton Of the euphausiid shrimps *Benthenphausia* and various species of *Thysanopoda Nematoscelis* and *Stylocheiron* have centers of abundance in the bathypelagic zone Deep water genera of mysids include *Gnathophausia Lophogaster* and *Eucopia* while the prawn families Hoplophoridae and Sergestidae have numerous bathypelagic representatives

5 Chaetognatha Certain species of arrowworms such as *Eukrohnia fowleri* and *Sagitta macrocephala* are predominantly bathypelagic

6 Echinodermata (Holothuroidea) The genera *Pelagothauria Enypniastes*, and *Galatheathuria* are bathypelagic the first two being medusalike in form

7 Protochordata (Thaliacea) While they are more abundant in the surface *layers the* salps doliolids and pyrosomes have been fished down to 3000 m

Deep-sea nekton This group consists largely of squids octopods and fishes The sperm whale also enters the deep sea where it finds some of *its squid* food

1 Moll . . .
squids (H
Chirotheu
portant part of the deep-sea nekton together with a few octopods such as *Cirrothauma, Amphitretus Vitreledonella* and *Vampyroteuthis* (Vampyromorpha)

2 Fishes Apart from a few squaloid sharks, the bathypelagic fish fauna consists of teleosts The most diverse groups are the stomiatoids (Isospondyli) with about 300 species, Myctophidae (lantern fishes Iniomi), about 250 species and the ceratioid angler fishes (Pediculati), about 90 species The few species forming the orders Lyomeri (gulper eels) Giganturoidea Cetunculi (whale fishes) and Miripinnati are entirely bathypelagic as are certain of the eels (Cyemidae Nemichthyidae) and Berycomorphi (for example Melamphaidae)

Distribution The bathypelagic fauna is most diverse in the tropical and temperate parts of the ocean Numerous species are found in all three temperature zones but many appear to have a more limited distribution

Each species also has a definite vertical occurrence Present findings suggest that there are three main vertical zones each with a characteristic community Here the term bathypelagic is used for the fauna between about 1000 and 2000 m that above (between 200 and 1000 m) being called mesopelagic and that below 2000 m abyssopelagic (Fig 1) The typical forms of the mesopelagic fauna (stomiatoids and lantern fishes) live in the twilight zone of the deep sea (between the 20 and 10°C isotherms) while the bathypelagic species (ceratioid angler fishes and *Vampyroteuthis*) occur in the dark cooler parts below the 10°C isotherm

Lastly numerous species of mesopelagic animals such as euphausiids prawns squids and fishes (particularly lantern fishes), undertake extensive diurnal vertical migrations moving upward into the productive surface layers to feed during the night Toward sunrise they begin to descend to their daytime levels *See* SCATTERING LAYER

Bioluminescence Perhaps the most conspicuous feature of pelagic deep-sea life is the widespread occurrence of luminescent species bearing definite light organs (photophores) Many of the squids and fishes have definite patterns of such lights as do some of the larger crustaceans (hoplophorid and sergestid prawns and euphausiids) Recent investigations (1958) with a bathyphotometer showed that flashes from luminescent organisms could be detected down to depths of 3750 m *See* BIOLUMINESCENCE, PHOTOPHORE GLAND

Deep sea benthic fauna There are two main ecological groups of bottom living animals (Fig 3) in the ocean organisms that attach to the bottom and those that freely move over the bottom

Attached benthic organisms This group consists of species that attach themselves to the sediments, rocks or to other organisms The more typical forms are included in the following list

1 Porifera Hexactinellida (glass sponges) about 375 species

2 Coelenterata Certain hydroids gorgonians pennatulids (sea fans) antipatharians (black cor

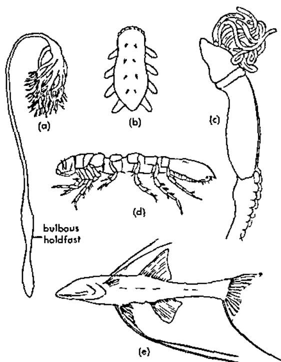


Fig 3 Benthic deep-sea animals (a) Sea pen showing bulbous holdfast *Umbellula* (b) Elasipod holothurian (sea cucumber) *Elpidia glacialis* (c) Head end and forepart of the trunk of a pogonophoran *Birsteinia vityasi* (d) Isopod crustacean *Macrostylus hadalis* (e) Bathypteroid fish *Benthosaurus*

als) actiniarians (sea anemones) and madrepore corals (*Lophohelia* and *Amphihelia*)

3 Crustacea Cirripedia (barnacles) such as *Scalpellum* and *Verruca* spp

4 Echinodermata Crinoidea (sea lilies) Numerous species of stalked crinoids live in the deep sea together with a number of unstalked forms

5 Protochordata Pogonophora (beard bearers) and certain ascidians (*Culeolous* spp)

Benthic crawlers and swimmers This group comprises the freely moving animals those that swim or crawl over the bottom or burrow into the sediments the upper layer of which has a rich bacterial flora

1 Annelida A few species of Polychaeta (bristle worms)

2 Gephyrea Certain species of echiuroid and sipunculoid worms

3 Crustacea In numbers of species and individuals the most important group of benthic deep sea crustaceans is the Peracarida represented by various species of cumaceans (*Bathycuma Macrocylindrus*) isopods (*Ischnomesus* and *Eurycope*) amphipods and tanaids (*Apseudes Neotanais*) Of the Eucarida the most prominent groups are the penaeid prawns and the Eryonidae There are also a number of crabs (*Platymaia Geryon Ethusa* and *Scyramathia*) and hermit crabs (numerous species of Axiidae)

4 Pycnogonida (sea spiders) Numerous species of the families Colossendeidae and Nymphonidae

5 Mollusca Certain of the Octopodidae (octopuses) and the cirromorph octopods live on the deep sea floor as do various gastropods scaphopods and lamellibranchs A small limpetlike mollusk (named *Neoplina galatheae*) was dredged the first time on the *Galathea* expedition in 1951 *See* MONOPLACOPHORA

6 Echinodermata These form an important part of the benthic fauna particularly the sea cucumbers (Holothuroidea) of the orders Elasipoda and Molpadonia Among the sea urchins (Echinoidea) the order Cidaroidea and the suborder Meridosternata mainly consist of deep sea species There are also various brittle stars (Ophiuroidea) and star fishes (Asteroidea)

7 Fishes The species of one group of cartilaginous fishes (Holocephali) live over the continental slope The main groups of benthic deep sea teleosts are the Bathypteroidae (Iniomi) Halosauridae and Notacanthidae (Heteromi) Macrouridae (rat tails) and Morinae (deep sea cods) (Anacanthini) Brotulidae Liparidae and Zoarcidae (Percomorphi)

Distribution The benthic fauna is most diverse in the temperate and tropical ocean although the arctic and antarctic areas have their characteristic species As in the pelagic fauna certain species occur in all three oceanic zones while others appear to have a more restricted occurrence

While a number of species—particularly among the polychaete worms gastropod mollusks and the brittle stars (Ophiuroidea)—range from littoral to abyssal regions most forms tend to live within smaller ranges of depth Present data suggest that there are typical communities of animals over the continental slopes (Fig 1) extending down to about 3000 m (bathyal zone) others occur below this in the abyssal zone Recent Danish and Russian exploration also suggests that the deep sea trenches (with depths over 7000 m) form another ecological zone (hadal zone) having certain characteristic species—those capable of living under pressures of 700 to 1000 atmospheres (barophilic species) This work also showed that life could exist at the very bottom of the ocean (down to depths of more than 10 000 m) and that species of certain groups such as sea anemones echiuroid and polychaete worms bivalves isopod and amphipod crustaceans sea cucumbers and Pogonophora occurred at depths beyond 9000 m

Lastly there is a decrease in the numbers of species and individuals with depth Russian biologists found that at depths of 8000–10 000 m the weight of animals per square meter of sea floor was about one fifth to one fifteenth the weight at depths of 1000–4000 m As the deep sea benthic fauna is dependent on organic matter originating in the upper plant bearing waters and as the amount reaching the bottom must decrease with depth the above findings are comprehensible It is also interesting that there are very few carnivorous animals such

as crabs brittle stars and starfishes below a depth of 7000 m It is the particle catchers such as the Pogonophora and ooze-eaters like sea cucumbers and echiuroid worms that make up most of the hadal fauna *See* MARINE ECOSYSTEM SEA WATER FERTILITY [N B M]

Bibliography A F Bruun Animals of the abyss *Sci American* 197(5) 50–57 1957 A F Bruun Deep sea and abyssal depths *Geol Soc Am Mem* 67 1 641 672 1957 *The Galathea Deep Sea Expedition 1950 1952* 1956 N B Marshall *Aspects of Deep Sea Biology* 1954

Deer

Any of a large number of ruminants of the family Cervidae found in North and South America and in Eurasia Male deer bear antlers which are replaced each spring females of the reindeer and caribou also members of the deer family have antlers



The white tailed deer *Odocoileus virginiana* length to 8 ft (From E L Palmer Fieldbook of Natural History McGraw Hill 1949)

The Virginia or white tailed deer *Odocoileus*

improvement and natural succession in cut-over forests and now is seriously overpopulated in many areas The mule deer *O hemionus* tends to replace the white tail west of the Great Plains Its population problems are similar to those of the white tailed deer *See* ARTIODACTYLA CARIBOU ELK MOOSE REINDEER [J D B]

Defecation

The process by which the fecal wastes that reach the lower colon and rectum are evacuated from the body Although the contents of the feces are complex and variable the body loses the following general types of material via this excretory route (1) small amounts of water (2) undigested and unabsorbed residues of ingested food (3) substances excreted from the body into the digestive tract particularly inorganic salts (4) compounds secreted into the intestine as part of the digestive process (5) living and dead microorganisms that normally inhabit the alimentary tract, and (6) products resulting from the chemical breakdown of the above mentioned materials *See* DIGESTIVE SYSTEM

The proximal colon which is largely under autonomous neural control receives the fecal mass from the small intestine and with several bursts of mass peristalsis during each 24-hour period drives its contents down to the lower colon and rectum Normally the fecal material is a liquid chyme on entry into the proximal colon It loses water to become a semisolid mass by the time it reaches the rectum This movement through the colon is augmented when food is ingested The distal colon which is considerably under control of the central nervous system can have its activity heightened by stressful environmental circumstances *See* NERVOUS SYSTEM

Except during infancy the act of defecation itself is under both reflex and voluntary control The need to defecate is signalled by autonomic afferent nerves when there is sufficient distention of the rectum This peripheral stimulus reflexly produces a contraction of the colon and a relaxation of the internal anal sphincter Subsequent to this reflex act voluntary relaxation of the external anal sphincter initiates evacuation of the rectum The voluntary component of defecation may also involve the diaphragm and abdominal musculature *See* BODY RHYTHM REFLEX CONDITIONED REFLEX UNCONDITIONED [R A M]

Definite composition, law of

This law states that a given chemical compound always contains the same elements in the same fixed proportions by weight Thus whatever its source silver chloride always contains 100 grams (g) of silver to every 32 85 g of chlorine If a compound is formed by the union of m atoms of one element each weighing a with n atoms of another element each weighing b the composition by weight of one molecule of the compound is in the ratio ma nb This must be the composition of any mass of the compound provided that all atoms of the same kind have the same weight It is now known that this is not usually the case but that the atoms of an element may consist of a number of isotopes having different masses However as long as any sample of the element always contains the same relative proportions of the isotopes the law still holds *See* ATOMIC WEIGHT ISOTOPE

This is not the case for lead Lead is the final product of the decay of three radioactive series the atomic weights from the three series being 206 from radium 208 from thorium and 207 from actinium Hence both the atomic weight of lead and the proportion of lead in its compounds will vary with the source of the lead *See* RADIOACTIVITY

Much more widespread and serious departures from the law of definite composition occur in a large variety of solid compounds (the non stoichi

ometric compounds) Non stoichiometry arises for a number of differing reasons In the silicate minerals such as olivine it occurs as a result of isomorphous replacement Thus olivine is $(Mg\,Fe)_2SiO_4$ and the proportion of magnesium to iron may vary widely from sample to sample Other solids are simply deficient in metal atoms Thus ferrous sul fide FeS rarely has an atomic ratio Fe/S of precisely unity Density measurements have shown that the lattice of the sulfur atoms remains intact but that some of the iron atoms are missing *See* CRYSTAL STRUCTURE NONSTOICHIOMETRIC COM POUNDS SILICATE MINERALS STOICHIOMETRY

[T C W]

Defoliant

A substance that can be used to remove leaves from plants without harming the remaining stems and fruits In modern agriculture defoliation is being used increasingly to hasten and facilitate harvest ing of crops

The most common agency of defoliation in na ture is frost frosted leaves dry up and fall off after a few days Some crops can be defoliated by graz ing animals for example some nurserymen turn sheep into their rose nurseries to get rid of the [illegible]

sult in defoliation For success however it was necessary that the plants be wet with dew and that the humidity remain high for a period of several days and nights

With the development of mechanical cotton pick ers defoliation is of great importance it allows the machines to get into the crop and harvest it as soon as the bolls are mature it removes the leaves that tend to get into the fingers of the picking de vices and stain the cotton green it reduces materially the manual labor of hand picking and in creases the efficiency of machine-picking

Because of the above advantages of defoliation much effort has gone into the search for other chemical defoliants New and improved formula tions containing calcium cyanamide have been produced borate chlorate combinations have proved to be good defoliants and 3 6 endoxohexa hydrophthalic acid (endothal) aminotriazole (ATA) ammonium thiocyanate and a number of other new chemicals are being used

True defoliation results from the formation of an abscission layer at the base of the petiole of the leaf Most of the above chemicals bring about this type of defoliation and the leaves abscise and drop from the plant Certain fortified oil emulsion con tact herbicides will kill plant leaves at a low ap plication cost but usually such killing does not result in abscission However if the cotton is heavily infested with bindweed Johnson grass or other heavy weed growth use of a contact herbicide will enable the grower to harvest the crop by machine whereas it would be impossible to harvest by ma chine or by hand if the weeds were alive Because modern cotton gins are able to separate this leafy trash from harvested cotton with a fair degree of success such herbicides are often used

Alfalfa clovers soybeans field beans and a good many flower and nursery crops are now defoliated Other crops such as rice and grain sorghums are sprayed with contact herbicides to prepare them for timely harvest The term preharvest desic cation is used to describe this process *See* AGRI CULTURE, AGRICULTURAL MACHINERY HERBICIDE PLANT GROWTH

[A S C]

Degaussing

Neutralization of the magnetization of a ship by properly located and oriented current carrying coils which produce a magnetic field of desired strength and direction

A steel ship has structural components of many different ferromagnetic characteristics Many parts become magnetized during the construction and retain that magnetization for a long time Other parts are soft iron and do not retain a magnetized con dition permanently but become magnetized by induction in the magnetic field of the earth The ship then has a magnetic field that consists of two parts the semipermanent component and the tem porarily induced component This magnetization of the ship causes deviation of the magnetic compass and may trigger magnetic mines or other explosive devices when the ship passes near them *See* MAG NETIZATION

One method for neutralizing the magnetic field of the ship is to install coils in which currents are maintained to produce components of the field that will neutralize the field due to the magnetization of the ship These coils are called degaussing coils

Sets of degaussing coils are arranged to compen sate separately for three components of the mag netization Since the largest component of the earth s field is vertical the main coils have their planes horizontal to produce a compensating verti cal field The two horizontal components are those parallel to and perpendicular to the length of the ship These components are compensated by sets of coils with vertical planes One set with planes perpendicular to the length of the ship compen sates for the fore and aft component A set with planes parallel to the length of the ship compen sates for the athwartship component

The current in the degaussing coils must be ad justable because of the variation in magnetization of the ship The induced component in particular varies as the ship changes position and direction As a ship moves south toward the magnetic equa tor the downward directed vertical component of the earth s field decreases and the horizontal com ponent increases South of the magnetic equator the vertical component is directed upward *See* GEOMAGNETISM

The horizontal component of the magnetization of the ship varies not only because of the change in magnitude and direction of the horizontal co ponent of the earth s field but also because of

changing orientation of the ship relative to the earth's field When the ship is parallel to the magnetic field of the earth the induced horizontal magnetization is entirely fore and aft For other orientations of the ship athwartship components appear [K V M]

Bibliography L B Loeb *Fundamentals of Electricity and Magnetism* 3d ed 1947

Degeneracy (quantum states)

A term referring to the fact that two or more stationary states of the same quantum mechanical system may have the same energy even though their wave functions are not the same In this case the common energy level of the stationary states is degenerate The statistical weight of the level is proportional to the order of degeneracy that is to the number of states with the same energy this number is predicted from Schrodinger's equation

Except for so called accidental degeneracy degeneracy is associated with special symmetries of the physical system and can be removed by destroying this symmetry For example an energy level of total angular momentum $\sqrt{J(J+1)}\ h\ 2-$ has a $(2J+1)$ fold degeneracy that results from the circumstance that the $2J+1$ allowed values of J_z the z component of total angular momentum all have the same energy In a magnetic field H along z this level splits into $2J+1$ energy levels since the energy now depends on J_z that is on the angle between H and the total angular momentum

In quantum mechanics and in other branches of mathematical physics the term degeneracy is employed also to characterize the eigenvalues of operators other than the energy operator See EIGENVALUE QUANTUM THEORY NONRELATIVISTIC [E C]

Degree of freedom (mechanics)

Any of the possible independent ways in which the space configuration of a mechanical system may change is called a degree of freedom of that system A material particle confined to a line in space can be displaced only along the line and therefore has one degree of freedom A particle confined to a surface can be displaced in two perpendicular directions and accordingly has two degrees of freedom A particle free in physical space has three degrees of freedom corresponding to three possible perpendicular displacements A system composed of two free particles has six degrees of freedom but a requirement that the particles remain a constant distance apart reduces the degrees of freedom to five two further requirements that each particle remain on a surface reduce the number of degrees to three

A requirement which reduces by one the degrees of freedom of a mechan al

time See CONSTRAINT

Any mechanical system may be conceived as a set of N particles in space subject to a certain number K of constraints reducing the number of degrees of freedom to $3N-K$ In the case of a rigid structure this number of degrees proves to be six regardless of the number of particles provided the number is greater than two [R A F]

Bibliography H C Corben and P Stehle *Classical Mechanics* 1950 J C Slater and N H Frank *Mechanics* 1947

Degree-day

An estimating unit for use in connection with fuel quantities for building heating A degree-day is based on the straight line relationship between fuel consumption and the extent to which the daily mean outside temperature falls below 65°F For example if the daily mean temperature is 52°F on a given day for that day 65 minus 52°F times 1 day equals 13 degree days For a specific building there will be twice as much fuel used on a day having 26 degree-days as during a day having 13 degree days If the mean daily temperature is above 65°F there are no degree-days that day

Once the fuel consumption for a specific building is known for a period for which the total number of degree-days is known the unit fuel consumption (fuel consumption per degree-day) can be determined Subsequent degree-day totals multiplied by this unit value determine when oil deliveries for example are to be made to that building before the tank is empty Another use of the degree-day is to reduce fuel consumption records for a building to a quantity per degree-day unit so that fuel data for different periods can be compared in other words the outside temperature variable is eliminated In short an important use of the degree day is as a guide to operating efficiency when current data are used The normal number of degree days for a locality is the average number to be expected in that locality for a given month or as a yearly total these figures can be used to predict fuel consumption [C ST]

Dehumidifier

Equipment designed to reduce the water vapor in the atmosphere

The atmosphere is a mechanical mixture of dry air and water vapor the amount of water vapor being limited by air temperature Water vapor is measured either in grains per pound of dry air or pounds per pound of dry air (7000 grains = 1 lb)

There are three methods by which water vapor may be removed (1) the use of sorbent materials (2) cooling to the required dew point and (3) compression with aftercooling

Sorbent type Sorbents are materials which are hygroscopic to water vapor, they are available in both solid and liquid forms Solid sorbents include silica gels activated alumina and aluminum bauxite Liquid sorbents include halogen salts such as lithium chloride lithium bromide and calcium chloride and organic liquids such as ethylene

diethylene and triethylene glycols and glycol derivatives

Solid sorbents may be used in static or dynamic dehumidifiers Bags of solid sorbent materials within packages of machine tools electronic equip ment and other valuable materials subject to moisture damage constitute static dehumidifiers An indicator chemical may be included to show by a change in color when the sorbent is saturated The sorbent then requires reactivation by heating at 300–350°F for 1–2 hours before reuse

A dynamic dehumidifier for solid sorbent con sists of a main circulating fan one or more beds of sorbent material reactivation air fan heater mechanism to change from dehumidifying to re-activation, and an aftercooler (Fig 1)

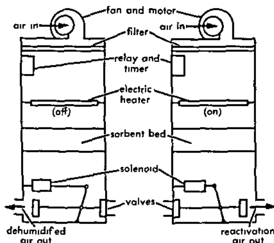


Fig 1 Diagrammatic arrangement of single-bed solid sorbent dehumidifier Dehumidifying cycle (left) and reactivation cycle (right)

A single bed dehumidifier operates on an inter mittent cycle of dehumidifying for 2 3 hours and then switches over to the reactivation cycle for 15–45 min No dehumidification can be obtained during the reactivation cycle A single bed unit is usually portable and is used for small storage rooms basement playrooms, home workshops and other small areas where moisture damage may be a problem The moist reactivation air is discharged to the outside

The dual bed machine is larger in capacity than the single bed unit and has the advantage of pro viding a continuous supply of dehumidified air (Fig 2) While one bed is dehumidifying the other bed is reactivating After a predetermined time interval the air cycle is switched to pass the air through the reactivated bed for dehumidification and to reactivate the saturated bed

The dew point of the effluent air of a fixed bed machine is lowest a the start of a cycle immedi ately after the reactivated bed has been placed in service (*see* DEW POINT) The dew point gradually rises as the bed absorbs the water vapor and even tually would be the same as the entering dew point when the vapor pressure of the sorbent reached the vapor pressure of the air and could no longer absorb moisture from the air The cut off point at which the absorbing bed is changed over to re-activation is fixed by the maximum allowable effluent dew point

A multibed unit with short operating cycles will reduce the range of effluent dew point to within a few degrees A unit with rotating cylindrical bed maintains a reasonably constant effluent dew point

The liquid sorbent dehumidifier consists of a main circulating fan sorbent air contactor sorbent pump and reactivator including contactor fan heater and cooler (Fig 3) This unit will control the effluent dew point at a constant level because dehumidification and reactivation are continuous operations with a small part of the sorbent con stantly bled off from the main circulating system and reactivated to the concentration required for the desired effluent dew point

Cooling type A system employing the use of cooling for dehumidifying consists of a circulating fan and cooling coil The coil may use a source of cold water obtained from wells or a refrigeration plant or may be a direct expansion refrigeration

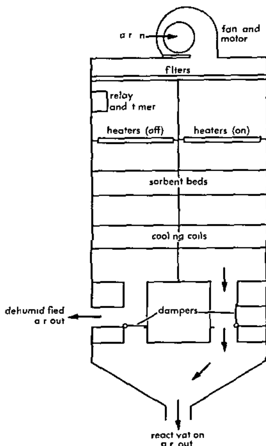


Fig 2 Diagrammatic arrangement of dual bed solid sorbent dehumid fier Air is being dehumidified through one bed (left) while the other bed (right) is being re activated

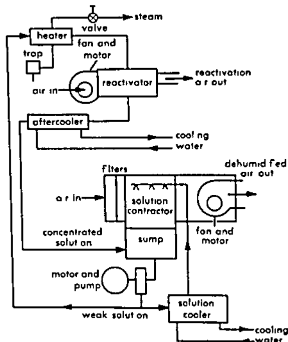


Fig 3 Diagrammatic arrangement of liquid sorbent dehumidifier with continuous dehumidifying and reactivation

coil In place of a coil a spray washer may be used in which the air passes through two or more banks of sprays of cold water or brine depending upon the dew point temperature required

When coils are used the leaving dew point is seldom below 35°F because of possible build up of ice on the coil When it is necessary to use coils for temperatures below 35°F, as in cold storage rooms either two coils are used so one can be defrosted while the other is in operation or only one coil is used and dehumidifying is stopped during the defrost period

A brine spray dehumidifier or brine-sprayed coil can produce dew point temperatures below 35°F

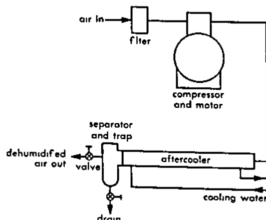


Fig 4 Diagrammatic arrangement of dehumidifying by compression and aftercooling

without frosting if properly operated and maintained

Compression type. Dehumidifying by compression and aftercooling is used when the reduction of water vapor in a compressed air system is required This is particularly important, for example, if the air is used for automatic control instruments or cleaning of delicate machined parts

If air is compressed and the heat of compression removed to bring the temperature of the air back to the temperature entering the compressor, condensation will take place and the remaining water vapor content will be directly proportional to the absolute pressure ratio of the compressed air (Fig 4)

For example if saturated air at 70°F (111 grains/lb of dry air) is compressed from atmospheric pressure (14 7 pounds per square inch absolute psia) to 88 psia (6 1 compression ratio), and cooled to 70°F the remaining water vapor in the compressed air will be 111/6 = 18 5 grains/lb of dry air If the air is expanded back to atmospheric pressure and 70°F, the dew point will be 24°F *See* PSYCHROMETRICS

The power required for compression systems is so high compared to power requirements for dehumidifying by either the sorbent or refrigeration method that the compression system is not an economical one if dehumidifying is the only end result required *See* DRYING. [J EV]

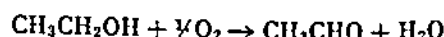
Bibliography American Society of Heating and Air Conditioning Engineers (ASHAE), *Heating, Ventilating and Air Conditioning Guide*, 1959. ASHAE, Symposium on dehumidification Journal section *Heating Piping Air Conditioning*, 29(4) 152-162, 1957, V R Dietz, *Bibliography of Solid Absorbents, 1943 to 1953*, Natl Bur Standards Circ 566 1956, J Everetts, Jr, Dehumidification methods and applications, *Heating Piping Air Conditioning*, 18(12) 121-124, 1946

Dehydrogenation

A reaction in which hydrogen is detached from its molecular linkage In a narrow sense it is the reverse of hydrogenation It is noteworthy, however that a major dehydrogenation process is based on the mildly destructive hydrogenation (reforming) of naphthas and naphthenes whereby aromatic compounds are formed From a strictly chemical view point some dehydrogenations may be viewed as a type of oxidation

Types. The several types of dehydrogenation reactions may be listed as follows

1 Vapor phase conversion of a primary alcohol to an aldehyde in the presence of a silver catalyst at 550°C



$$CH_3CH_2OH + \tfrac{1}{2}O_2 \rightarrow CH_3CHO + H_2O$$

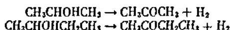
2 Dehydrogenation of a side chain as in the preparation of styrene from ethylbenzene in the presence of a promoted zinc oxide catalyst



$$C_6H_5C_2H_5 \rightarrow C_6H_5CH{=}CH_2 + H_2$$

The reaction is endothermic and hot flue gases are used to maintain a reaction temperature of 600°C

3 Vapor phase dehydrogenation of isoalcohols as in the preparation of acetone from isopropyl al cohol and methyl ethyl ketone from secondary bu tanol



$$CH_3CHOHCH_3 \rightarrow CH_3COCH_3 + H_2$$
$$CH_3CHOHCH_2CH_3 \rightarrow CH_3COCH_2CH_3 + H_2$$

4 Catalytic reforming of naphthas and naph thenes in the presence of a platinum catalyst for the production of aromatics for high octane gaso line, toluene for TNT and ortho , meta , and para xylenes for oxidation to the corresponding phthalic acids Catalytic reforming is usually carried out by feeding a naphtha (after pretreating with hydrogen to remove catalyst poisons) and hydrogen mixture to a furnace where the mixture is heated to about 450–520°C and then passed through a series of fixed bed catalytic reactors at hydrogen pressures of 100–1000 psi

All four of the listed types of dehydrogenation are of major industrial importance They account for the production of billions of pounds of organic compounds that enter into the manufacture of lu bricants explosives plastics plasticizers and elas tomers

Thermodynamics. Considering the close relation ship between hydrogenation and oxidation which are generally exothermic reactions it would be ex pected that unmodified dehydrogenations would be endothermic The dehydrogenation of isoalcohols and ethylbenzene are indeed endothermic reactions that require the application of much heat In the dehydrogenation of alcohols to aldehydes a heat balance can be achieved by using a hybrid reac tion involving partial oxidation *See* HYDROGENA TION, UNIT PROCESSES [P H G]

Delay line

A transmission line (as nearly dissipationless as possible) or an electric network approximation of it which if terminated in its characteristic imped ance will reproduce at its output a waveform ap plied to its input terminals with little distortion but at a time delayed by an amount dependent upon the electrical length of the line

If a transmission line is dissipationless which will be the case if its series resistance is zero and its shunt conductance is also zero it will have a characteristic impedance of

$$Z_0 = \sqrt{L/C} \qquad (1)$$

where L is the series inductance and C the shunt capacitance per unit length of the line *See* TRANS MISSION LINES

The velocity of propagation of a signal along the line is

$$v = \frac{1}{\sqrt{LC}} \qquad (2)$$

Therefore the time required for the pulse to pro pagate a distance x along the line is

$$T_d = x\sqrt{LC} \qquad (3)$$

Such a line, terminated in its characteristic imped ance, is shown in Fig 1, and as indicated the out put pulses reproduce the input at a delayed time T_d

The lumped circuit approximation of the trans mission line is shown in Fig 2 If the inductance and capacitance per section are L_1 and C_1 then the total time delay is

$$T_d = n\sqrt{L_1C_1} \qquad (4)$$

where n is the number of sections

If the delay line is not terminated in its charac teristic impedance there is multiple reflection back and forth along the line For example if the receiv ing end of the line is unterminated as in Fig 3

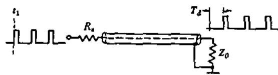


Fig 1 Transmission line as delay line



Fig 2 Lumped-circuit delay line

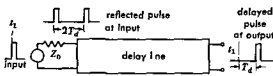


Fig 3 Reflection due to unterminated receiving end

but the sending end is terminated in its character istic impedance the receiving and reflection coeffi cient is positive and a delayed pulse appears at the output as given by Eq (3) and (4) In addi tion a pulse of the same polarity appears at the in put terminals delayed by twice this amount (the time required for a pulse to travel to the end and back) If the receiving end is terminated in a short circuit as shown in Fig 4 the receiving end reflec tion coefficient is negative and no pulse appears at the output although an inverted pulse will appear at the input terminals with the same time delay as before

Various applications make use of the short cir cuit and open circuit of delay lines including line

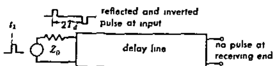


Fig 4 Reflection due to short-circuited receiv

controlled pulse generators (see BLOCKING OSCILLATOR, PULSE GENERATOR)

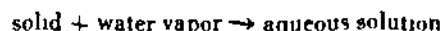
Delay lines are also used for establishing a time sequence for the occurrence of events. A delay line with a total length equal to the greatest time delay required in a system may be used as a basic element. Pulses occurring at intermediate times may be obtained from taps at various points along the line. A specific application is found in the synchronizing signal generator of the television system. Also the lumped circuit delay line is an essential element of the wide band distributed amplifier. See TIME DELAY CIRCUITS. [G M C]

Bibliography G M Glasford *Fundamentals of Television Engineering*, 1955; J Millman and H Taub *Pulse and Digital Circuits* 1956

Deliquescence

The absorption of atmospheric water vapor by a crystalline solid until the crystal eventually dissolves into a saturated solution. This behavior is well known for certain salts such as hydrated calcium chloride, $CaCl_2 \cdot 6H_2O$, and zinc chloride, $ZnCl_2$, but it is a property of all soluble salts in air of sufficiently high humidity.

Thermodynamically the condition for deliquescence is that the partial pressure of the water vapor in the air exceed the vapor pressure (aqueous tension) of the water in the saturated solution of the salt. Then the reaction



solid + water vapor → aqueous solution

will occur spontaneously. The speed at which the process takes place depends upon the rate of diffusion of water vapor into the crystal lattice, crystal size, and other factors. The process will stop when the water vapor in the atmosphere is depleted to the point at which its partial pressure equals that of the saturated solution.

In general, substances which are highly soluble in water have a greater tendency to deliquesce since concentrated solutions will have a lower vapor pressure. At 25°C the vapor pressure of pure water is 23.8 mm Hg, whereas that of a saturated solution of $CaCl_2 \cdot 6H_2O$ is only 7.0 mm Hg; this salt then will deliquesce in an atmosphere where the relative humidity exceeds 7.0/23.8 or 30%. For $ZnCl_2$ the situation is much more extreme: it will deliquesce at a relative humidity of 10%. Ordinary sugar (sucrose) at 25°C will deliquesce at humidities above 85%.

Crystalline solids also may absorb water by increasing their water of hydration if the dissociation pressure of the hydrated species to be formed is less than the partial pressure of the water vapor. It is this process, not deliquescence, which is the opposite of efflorescence.

Deliquescent substances can be used to remove water vapor from air, although they have no special advantage over substances which merely add water of hydration and remain crystalline. See DESICCANT; DRYING; EFFLORESCENCE; SOLUTION; VAPOR PRESSURE. [R L S]

Delta

A deposit built by jetlike flow of sediment-laden water into or within a permanent body of water. Herodotus (fifth century B.C.) used the term for delineating the flat alluvial plain inclosed by the branching mouths of the Nile River because this region had the shape of the Greek letter Δ (delta). Geologists now use the term to include the entire deposit formed not only above but also below mean water level near mouths of rivers, tidal inlets, submarine canyons, and storm-induced washovers of barrier islands. See FLOOD PLAINS; PLAINS; SHORE PROCESSES.

Delta formation and growth. The manner in which the bed load and suspended sediment of a stream is deposited upon entering a basin depends in part upon how the inflow decelerates. If there is little density contrast between inflowing and entraining fluids, forward flow becomes negligible at about a tenth of the distance observed for such a condition when a strong density contrast occurs. This former case normally happens when a river enters a lake or shallow marine embayment, and most of the sediment is dropped quickly in a Gilbert-type delta consisting of top-, fore-, and bottom-set beds (Fig. 1). Should there be a pronounced density contrast, as when lighter river water enters the sea, typical marine deltas form.

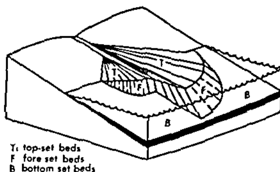


Fig. 1 Schematic block diagram showing two stages of growth of a typical Gilbert-type delta. (Adapted from P. H. Kuenen, Marine Geology, Wiley, 1950)

Should the inflow be the denser, as in the case of very cold muddy water entering a warm lake or a density current flowing basinward through a submarine canyon, such inflow seeks an appropriate density level, often the bottom, and much of the deposition occurs a considerable distance away from the orifice of water discharge. Thus, deposits of this type in Lake Mead reached a thickness of 135 ft against Hoover Dam within 11 years even though the Colorado River empties into the head of the reservoir 75 miles away. Similarly, of the 44,000 sq mi of deltaic deposits formed by the Mississippi River during the past 1,000,000 years, 11,000 sq mi occur as subaerial deltaic plain, 17,500 occur on the continental shelf, 8500 occur on the inner continental slope, and 7000 occur as a bulge



Fig 2 Changes in cuspate deltaic deposit formed at mouth of New Brazos River Texas (a) New bar in front of original bar and cuspate delta on December 18 1938 (b) Remodeled delta deposit on October 31 1949 (*Aerial mosaics by Jack Ammann Inc*)

in the submarine contours below depths of 1000 fathoms

Characteristics of modern deltas While over 150 major deltas are forming today not all streams have deltas Of rivers over 100 miles in length along the southern Baltic Sea coast 48 have deltas and 26 do not Even the mighty Amazon River empties into a tidal estuary Typical statistics pertaining to some modern deltas are given in the accompanying table

Delta thicknesses vary widely The Nile River is depositing a silt layer over the sandy bottom of a marine embayment 50 ft in depth In contrast the Mississippi River is building into deep water and drill holes 20 miles offshore show 850 ft of recent deltaic deposits which are known to be still thick

Statistics pertaining to modern deltas

River	Dimension of subaerial delta statute miles		Amount of sediment discharged		Annual extension of subaerial delta	
	Length	Breadth	River water by weight (avg) parts per million	Annual volume of sediment cubic miles	Period of measurement years	Approximate distance ft
Mississippi s present birdfoot delta	12	30	550	0 068	1838 1947	250
Hwang Ho	300	470*	50 000†		1870–1937	950
Ganges-Brahmaputra	220	200	870	0 043 (Ganges only)		
Rhone into Mediterranean Sea	30	47	400–590	0 005	1737 1870	190
Danube	46	46	310	0 008		40
Nile (prior to barrages)	96	145	1600	0 001	1100–1870	45
Colorado above Hoover Dam	43	0 05–0 6	8300	0 032	1936–1948	3 6 miles (gorge)
Euphrates-Tigris	350	90			1793 1853	180

* Includes 100 miles of nondeltaic Shantung ula † Maximum is 400 000 ppm

ening seaward The weight of such new deposits on underlying ductile prodelta clays sometimes causes mudlumps to form Off the Mississippi River passes such mudlumps may temporarily reach a height of 10 ft above sea level achieve an areal extent of 40 acres and have vents and fissures discharging gas mud or both *See* Gulf of Mexico

Pattern of deposition The position shape and extent of a delta vary widely in time Major distributaries are subject to bifurcation abandonment or sudden relegation to a minor role as stream diversions take place locally or upstream Such changes are most common when a delta builds outward a considerable distance and the stream finds a shorter route to the sea or lake during flood During the past 5000 years the Mississippi River has formed five clearly discernible lobate delta complexes over an area extending 100 miles east and west of New Orleans A sixth major shift is imminent by 1975 unless engineering works are constructed at Old River 300 miles above the present passes for the outflow of the Atchafalaya River has increased its share from 5 to 24% of the Mississippi River's water at this diversion point between 1880 and 1950

The changing balance between the forces of erosion and accretion off river mouths is typified by the history of the Brazos River Texas after the U S Corps of Engineers shifted the river to a new and leveed channel in 1929 A transverse bar formed off the river mouth by July 1931 By October 1934 the river cut a channel through the deposit and by December 1938 a new transverse bar formed in front of the original bar (Fig 2*a*) During a period of relative stability a bifurcation of the channel occurred at the second bar site and a relatively permanent triangular shaped deposit was formed by 1944 However the increased wave action during the hurricanes of 1945 and 1949 again remolded a cuspate delta (Fig 2*b*)

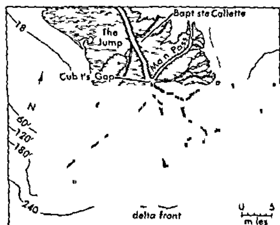


Fig 3 Features of modern Mississippi birdfoot delta (From H N Fisk E McFarland Jr C R Kolb and L J Wilbert Sedimentary framework of the modern Mississippi delta J Sediment Petrol 24 76–99 1954)

Coastal outlines of deltas are often other than the traditional triangle depending upon the balance achieved between rates of deposition coastal erosion and local and regional subsidence Typical shapes include the digitate or birdfoot (lower 15 miles of present Mississippi River delta as shown in Fig 3) lobate with branching distributaries (Mississippi River delta below Old River) arcuate with branching distributaries (Niger delta) cuspate with single outlet and flank depressions (Brazos delta) and multilobate with funnel shaped distributaries kept open by tidal scour (Ganges Brahmaputra)

Engineering problems Despite many difficult engineering problems such famous cities as Calcutta Shanghai Venice Alexandria (Egypt) and New Orleans have been constructed on deltas These problems include continually shifting and extending shipping channels a lack of firm footing except on narrow natural levees steady subsidence which may reach a rate of as much as 5 ft per century and poor drainage because of extremely flat slopes which may be no greater than 1/10 ft per mile Such flat slopes permit extensive flooding during high river stages Submergence to a depth of 15 ft or more during typhoons or hurricanes with high loss of life is not uncommon [C C B]

Deltaic sediments Deltaic sediments predominantly consist of the products of weathering and erosion from the drainage area of the river Subordinately other sediments may occur as for example salt gypsum and anhydrite formed in dry climates by evaporation in shallow lagoons or peat developed in swamps in humid areas

The study of deltaic sediments is of great importance for the understanding of the deltaic series of the geologic past Ancient deltaic sediments have been found to contain large quantities of petroleum thus an understanding of deltaic sedimentation is essential in oil exploration

Streams carry the finer part of their sediment load (silts and clays) in suspension the coarser part (gravel and sand) is transported along the bottom as bed load The material is distributed over the delta in four different ways (1) channel and bank sediments consisting of the coarsest material available (2) flood plain deposits of fine material carried in suspension and deposited in the low areas between river courses (3) delta front littoral barrier island and dune deposits resulting from the winnowing and redistribution of bed load material by waves currents and winds and (4) offshore prodelta and delta front sediments consisting of silts and clays distributed by currents over the sea floor adjacent to the delta Both the flood plain and offshore deposits grade from silts and silty muds close to the river to heavy clays at greater distances

Continued deposition will raise the surface of the delta and finally the river will change its course and start developing a new delta in an adjacent lower part of the basin Marine erosion and winnowing of the old abandoned area then result in

the formation of a submerged sheet of sand which occupies the position of the outer part of the former delta

In the zone of active deposition deltaic sediments can build up rapidly at rates of up to 1 ft per year Consequently, in a rapidly subsiding basin the total amount of deltaic sediments accumulated may be great In the modern Mississippi birdfoot delta (Fig 3) an estimated 200 000,000 tons of sediment have accumulated in the last 450 years During the same time, the rate of sedimentation in the entire Mississippi delta was 560,000,000 tons per year

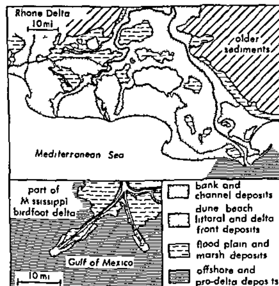


Fig 4 Examples of deltaic sediment distribution

Basically a delta consists of a framework of elongated bodies of coarse material formed as littoral and channel deposits with a matrix of fine grained flood plain marsh and offshore sediments (Fig 4) The ratio between framework and matrix depends on the relative amounts of coarse and fine material in the sediment load of the river and upon the vigor and efficiency of the marine reworking processes Deltas formed by rivers which predominantly carry sand and gravel or deltas formed in basins with strong wave action are largely sandy (Rhone delta), deltas of large rivers with long lower courses and deltas formed in quiet basins consist predominantly of clayey flood plain and marsh sediments (Mississippi delta) *See* ESTUARINE OCEANOGRAPHY, MARINE SEDIMENTS, SEDIMENTATION (GEOLOGY) *see also* MARINE MARSH

[T V A]

Bibliography C C Bates Rational theory of delta formation, *Bull Am Assoc Petrol Geologists,* 37(9) 2119–2162, 1953 H N Fisk, E McFarland Jr, C R Kolb and L J Wilbert, Sedimentary framework of the modern Mississippi delta *J Sediment Petrol*, 24 76–99, 1954, H N Fisk and E McFarland, Jr, Late quaternary deltaic deposits of the Mississippi River, in A Poldervaart (ed), *Crust of the Earth,* Geol Soc Am Spec Paper 62, 1955, C R Kolb and J R Van Lopik, *Geology of the Mississippi River Deltaic Plain Southeastern Louisiana,* U S Army Engineer Waterways Expt Sta Tech Rept 3-483 1958, C Kruit, Recent sediments of the Rhone delta, *Verhandel Ned Geol Mijnbouwk Genootschap,* 15 357–514

Delusion

A conviction based upon faulty perceptions feel

differentiate delusions from fixed faulty, and rigid ideas particularly if they are shared by small and large groups

Delusions seen in psychotic patients are usually classified according to content The main groups are as follows (1) persecutory or paranoid delusions these may be systematized and have a logic of their own as in paranoia, or may be unsystematic and incoherent, (2) delusions of influence related to paranoid delusions one of the most frequent forms is the delusion of being hypnotized or under a spell (3) delusions of grandeur in which the patient imagines he is an extraordinary person or that he has extraordinary powers, some historical figures had both grandiose and paranoid delusions for example Nero Herodotus George II and probably Hitler and Stalin (4) self depreciatory delusions occurring primarily in depressed patients these patients have an exceedingly low self esteem

gross abnormalities and diseases, these are also referred to as hypochondriacal delusions *See* PARANOIA, SOMATIZATION, *see also* ABNORMAL BEHAVIOR

Most delusions are products of a regression to primitive or primary thinking in which logic and the sense of reality are not yet established They are encountered in all types of psychoses particularly in schizophrenias and organic deficit states, and in many forms of intoxications and toxic psychoses The most important dynamic mechanism underlying delusions is projection Delusions can be transient or chronic, stationary or progressive Treatment varies depending on the therapy of the underlying disorder *See* PSYCHOSIS

[F C R]

Bibliography N Cameron and A M Cameron *Behavior Pathology,* 1951

ening seaward The weight of such new deposits on underlying ductile prodelta clays sometimes causes mudlumps to form Off the Mississippi River passes such mudlumps may temporarily reach a height of 10 ft above sea level achieve an areal extent of 40 acres and have vents and fissures discharging gas mud or both *See* GULF OF MEXICO

Pattern of deposition The position shape and extent of a delta vary widely in time Major distributaries are subject to bifurcation abandonment or sudden relegation to a minor role as stream diversions take place locally or upstream Such changes are most common when a delta builds outward a considerable distance and the stream finds a shorter route to the sea or lake during flood During the past 5000 years the Mississippi River has formed five clearly discernible lobate delta complexes over an area extending 100 miles east and west of New Orleans A sixth major shift is imminent by 1975 unless engineering works are constructed at Old River 300 miles above the present passes for the outflow of the Atchafalaya River has increased its share from 5 to 24% of the Mississippi River s water at this diversion point between 1880 and 1950

The changing balance between the forces of erosion and accretion off river mouths is typified by the history of the Brazos River Texas after the U S Corps of Engineers shifted the river to a new and leveed channel in 1929 A transverse bar formed off the river mouth by July 1931 By October 1934 the river cut a channel through the deposit and by December 1938 a new transverse bar formed in front of the original bar (Fig 2*a*) During a period of relative stability a bifurcation of the channel occurred at the second bar site and a relatively permanent triangular shaped deposit was formed by 1944 However the increased wave action during the hurricanes of 1945 and 1949 again remolded a cuspate delta (Fig 2*b*)

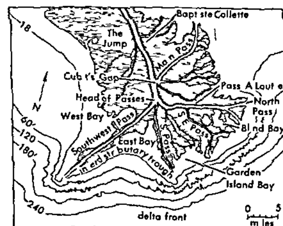


Fig 3 Features of modern Mississippi birdfoot delta (From H N Fisk E McFarland Jr C R Kolb and L J Wilbert Sedimentary framework of the modern Mississippi delta J Sediment Petrol 24 76–99 1954)

Coastal outlines of deltas are often other than the traditional triangle depending upon the balance achieved between rates of deposition coastal erosion and local and regional subsidence Typical shapes include the digitate or birdfoot (lower 15 miles of present Mississippi River delta as shown in Fig 3) lobate with branching distributaries (Mississippi River delta below Old River) arcuate with branching distributaries (Niger delta) cuspate with single outlet and flank depressions (Brazos delta) and multilobate with funnel shaped distributaries kept open by tidal scour (Ganges Brahmaputra)

Engineering problems Despite many difficult engineering problems such famous cities as Calcutta Shanghai Venice Alexandria (Egypt) and New Orleans have been constructed on deltas These problems include continually shifting and extending shipping channels a lack of firm footing except on narrow natural levees steady subsidence which may reach a rate of as much as 5 ft per century and poor drainage because of extremely flat slopes which may be no greater than ½₀ ft per mile Such flat slopes permit extensive flooding during high river stages Submergence to a depth of 15 ft or more during typhoons or hurricanes with high loss of life is not uncommon [C C B]

Deltaic sediments Deltaic sediments predominantly consist of the products of weathering and erosion from the drainage area of the river Subordinately other sediments may occur as for example salt gypsum and anhydrite formed in dry climates by evaporation in shallow lagoons or peat developed in swamps in humid areas

The study of deltaic sediments is of great importance for the understanding of the deltaic series of the geologic past Ancient deltaic sediments have been found to contain large quantities of petroleum thus an understanding of deltaic sedimentation is essential in oil exploration

Streams carry the finer part of their sediment load (silts and clays) in suspension the coarser part (gravel and sand) is transported along the bottom as bed load The material is distributed over the delta in four different ways (1) channel and bank sediments consisting of the coarsest material available (2) flood plain deposits of fine material carried in suspension and deposited in the low areas between river courses (3) delta front littoral barrier island and dune deposits resulting from the winnowing and redistribution of bed load material by waves currents and winds and (4) offshore prodelta and delta front sediments consisting of silts and clays distributed by currents over the sea floor adjacent to the delta Both the flood plain and offshore deposits grade from silts and silty muds close to the river to heavy clays at greater distances

Continued deposition will raise the surface of the delta and finally the river will change its course and start developing a new delta in an adjacent lower part of the basin Marine erosion and winnowing of the old abandoned area then result in

An important genus with aleuriospores is *Papularia* (*Coniosporium*) *Papularia* has 1 celled aleuriospores which are dark and lens shaped (lenticular) *P sphaerosperma* is found on the weeds *Phragmites*

An important genus with meristem phialospores is *Phialophora* *Phialophora* have simple or branched phialophores bearing several frequently flask shaped phialides The phialospores are one celled *P cinerescens* causes a vascular wilt of carnations

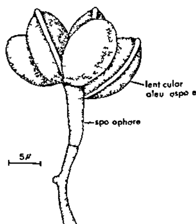


Fig 3 *Papularia sphaerosperma (Coniosporium arundinis)* Sporophore short bearing a cluster of 1-celled lenticular dark aleuriospores *(After G Goidanich 1938)*

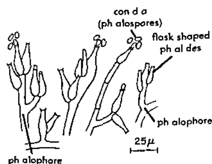


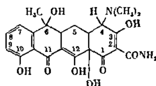
Fig 4 *Phialophora fastigiata* Phialophore with clusters of flask-shaped phialides *(After F N van Beyma thoe Kingma 1943)*

Important genera with terminus phialospores are *Helminthosporium* and *Cercosporella* *Helminthosporium* have conidiophores which are more or less irregular or bent and bear conidia successively on new growing tips The conidia are dark cylindrical with rounded ends and several celled The 175 parasitic species frequently cause leaf spots of grasses and are stages of *Pyrenophora* or *Cochiobolus* *H sativum* attacks wheat barley and rye *Cercospora* have conidiophores similar to those of *Helminthosporium* The conidia are threadlike (filiform) and multiseptate The genus differs from *Cercosporella* only in the dark colored conidiophores There are 400 parasitic species chiefly causing leaf spotting *C beticola* is found on sugar beet and *C apii* on celery [N F B]

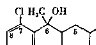
Demethylchlortetracycline

A broad spectrum antibiotic (trademark Declomycin) introduced by Lederle Laboratories in 1959 Its isolation and identification were reported by J R D McCormick and coworkers in 1957 It was the fourth tetracycline drug to be made available for clinical therapy The in vitro activity of the new compound falls within the same bacterial range as that of previous tetracycline antibiotics Studies have shown that demethylchlortetracycline demonstrates greater antibacterial activity per unit weight than previous tetracyclines and that its antibacterial activitity in the blood is markedly more prolonged than that of tetracycline chlortetracycline or oxytetracycline *See* CHLORTETRACYCLINE OXYTETRACYCLINE TETRACYCLINE

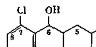
Structure Structure studies have shown that demethylchlortetracycline is designated as 4-dimethylamino 1 4 4*a* 5 5*a* 6 11 12*a* octahydro 3 6 10 12 12*a* pentahydroxy 1 11 dioxo-2 naphthacenecarboxamide The new compound differs from chlortetracycline by the absence of a methyl group in the 6 position of the basic tetracycline molecule The structural relationships of the four tetracyclines are shown in the formulas



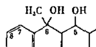
Tetracycline



Chlortetracycline



Demethylchlortetracycline



Oxytetracycline

Industrial production Demethylchlortetracycline is produced by a mutant strain of *Streptomyces aureofaciens* Duggar the microorganism which originally produced chlortetracycline The mutant strains producing demethylchlortetracycline are usually characterized by reddish brown pigmentation Mutations were achieved spontaneously and by treatment with mutagenic agents such as ultraviolet irradiation and nitrogen mustard

The new antibiotic is produced by a method similar to that used for the production of chlortetracycline A suitable fermentation medium may be prepared with the following substances starch lard oil $(NH_4)_2SO_4$ $CaCO_3$ trace elements and corn steep liquor

Comparison with tetracyclines One of the advantages of demethylchlortetracycline over p

ous tetracyclines is its increased stability in acids and alkali Chlortetracycline in an aqueous solution with a sodium carbonate buffer at pH 9 85 has a half life of 29 2 min at 23°C. By contrast demethyl chlortetracycline loses no more than 6% of its activity in 24 hours under the same conditions Tetracycline is completely destroyed in less than 5 min in 1 N sulfuric acid at 100°C. On the other hand demethylchlortetracycline loses only about 2% of its activity in 1 N sulfuric acid at 100°C after 15 min

M Finland and coworkers at Harvard University have established in studies among healthy volunteers that demethylchlortetracycline produces much higher and longer sustained levels of antibacterial activity in the blood serum after oral administration than tetracycline chlortetracycline or oxytetracycline The group also found that the biological half life of the new compound was 44% higher than that of tetracycline whereas its rate of renal clearance was about 43% of that of the older drug

The Harvard group also has reported a laboratory comparison of the sensitivity of 861 bacterial strains to demethylchlortetracycline and to tetracycline Of these strains 680 were regarded as susceptible to antibiotic therapy The new compound was found to be more effective in 62% of these strains tetracycline was more effective in 10% and no difference was noted in 28% Tetracycline showed more activity against 48% of 181 strains which are not generally regarded as clinically susceptible to drugs of this type Demethylchlortetracycline was more active in 16% of these strains and no difference was noted in 36% Of all of the 861 strains both sensitive and resistant to tetracyclines demethylchlortetracycline showed more effectiveness than tetracycline against 52% Tetracycline was more active against 18% whereas 30% of the strains were equally susceptible to both drugs

Clinical use Demethylchlortetracycline has a wide range of clinical usefulness corresponding generally to the range of usefulness of the other tetracyclines It is rapidly absorbed from the intestinal tract and is distributed into all body tissues Average daily dosage for adults is 600 mg divided into four doses For children the average dose is approximately 3 mg per pound of body weight per day

Clinical investigators have reported on the use of the drug in the treatment of a wide range of diseases including pneumonia genitourinary infections brucellosis acute childhood infections pustular dermatoses gonorrhea lymphogranuloma venereum and granuloma inguinale Studies indicate that the drug is as effective as or more effective than tetracycline in doses approximately 60% of

Untoward side effects of the new compound have not been found to be frequent or serious Among those reported have been nausea vomiting loose stools and some dermatologic effects including dermatitis pruritis and photosensitivity Side effects rarely have caused treatment to be discontinued Some investigators have reported that untoward reactions can be expected less frequently with demethylchlortetracycline than with tetracycline because of the smaller average dose of the new compound administered [J R.D M]

Demodulator

The stage in a radio television radar, or other receiver at which demodulation of the received signal takes place Thus in a tuned radio-frequency (rf) receiver the demodulator separates the audio-frequency (af) signal from the amplified incoming rf carrier signal A demodulator is often called a detector In a superheterodyne receiver the af signal is separated from the carrier signal at the second detector because the converter or first detector merely serves to change the modulated rf carrier signal to a modulated intermediate-frequency carrier signal In a frequency modulation receiver the demodulator converts carrier frequency changes into corresponding audio signals In a color television receiver the demodulator extracts color primary for color difference signals from the incoming modulated carrier signal *See* DETECTOR [J MR]

Demospongiae

A class of the phylum Porifera including sponges with a skeleton of one- to four rayed siliceous spicules or of spongin fibers or both Several genera lack a skeleton and it is through a study of these seemingly primitive forms that the complicated structure of most adult Demospongiae may be understood The Demospongiae constitute the most abundant and widely distributed group of sponges occurring in the sea from the tidal zone down to abyssal depths (at least to 5500 meters) One family has invaded fresh water The species vary in size from thin encrustations several centimeters in diameter to huge cake shaped forms which may measure up to 2 meters in diameter

Comparative studies of the embryology and early attached stages of sponges of the class Demospongiae suggest at least two evolutionary lines within this group One line of development the subclass Ceractinomorpha is characterized by the presence of incubated stereogastrula larvae the other the subclass Tetractinomorpha includes oviparous species with stereogastrula larvae However the more primitive families have incubated amphiblastula like larvae

Ceractinomorpha Among the Ceractinomorpha the genus *Halisarca* lacking skeletal elements is a primitive form The larva of *Halisarca* is a stereogastrula or parenchymula with an outer layer of flagellated cells and an inner mass of presumptive

ectomesenchymal cells The outer flagellated cells lose their flagella migrate into the interior and later differentiate into choanocytes Other cell types characteristic of the adult sponge differentiate and inhalant canals begin to form The young sponge soon develops a single internal cavity lined with choanocytes and an oscular opening breaks through at the apex At this stage of development called the rhagon stage the young *Halisarca* is essentially identical with the asconoid grade of construction found in some Calcarea Later folds in the choanocytal layer lead to the formation of flagellated chambers and a transitory syconoid grade of construction exists Eventually the flagellated chambers are isolated from one another through the appearance of exhalant canals which converge on the oscula The adult sponge has a simple leuconoid structure with elongate flagellated chambers having wide openings into exhalant canals and communicating with the surface pores by means of an inhalant canal system *Aplysilla* a closely related genus of sponges with a branching nonanastomosing fibrous skeleton has a similar developmental history except that the earliest rhagon stage is syconoid in structure with a folded choanocytal layer *See* CALCAREA

Metamorphosis from the larval condition in the Demospongiae characteristically involves a transitory rhagon stage with a simple leuconoid canal system During further development the flagellated chambers become isolated between the inhalant and exhalant canals to produce leuconoid canal systems of varying grades of complexity

In form ceractinomorph sponges vary from encrustations thin or massive to lobate and upright branching colonies The shallow water species tend to be more plastic in form than deep water species which usually exhibit little intraspecific variation in shape

Tetractinomorpha In the other subclass of Demospongiae the Tetractinomorpha the skeletonless genus *Oscarella* has a primitive structure Cleavage results in a solid mass of cells (morula) which later becomes hollow by cytolysis of the interior cells rich in food reserves Upon being freed from the parent the hollow larva is made up of a single layer of flagellated cells and is known as an amphiblastula After a short free swimming period the larva attaches to the substrate by its anterior pole and flattens out Gastrulation occurs at this point as the anterior half of the larva invaginates The blastopore closes as the edges of the now double-layered larva push together against the substratum The internal layer of cells is now thrown into folds which pinch off to form cavities which will become the flagellated chambers Simultaneously a depression forms in the apical ectoderm and the rudiment of the exhalant canal system appears The depression deepens to form a cavity giving off branches which push their way among the flagellated chambers The latter finally open into these cavities which become the exhalant canals

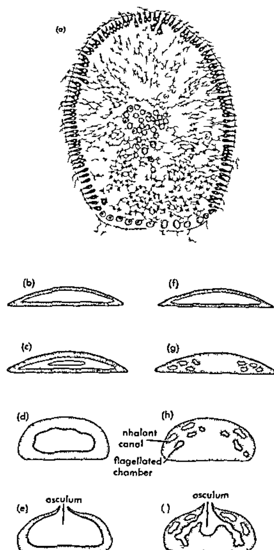


Fig 1 (a) Free-swimming parenchymula larva of *Halisarca* (b e) Metamorphosis of *Halisarca* (b) Newly settled parenchymula external flagellated cells migrate internally (c) internal cavity appears (d) choanocytes line the central cavity (e) osculum breaks through and young asconoid stage is formed (f) Metamorphosis of *Aplysilla* (f) newly settled parenchymula external flagellated cells migrate internally (g) islands of choanocytes form in the internal mass of cells (h) flagellated chambers and inhalant canals appear ( ) flagellated chambers join exhalant canal system which opens through osculum young syconoid sponge is formed (*After Lévi 1956*)

inhalant canals push in from the surface of the sponge and join the flagellated chambers The metamorphosed larva of *Oscarella* assumes the leuconoid grade of construction from the start with isolated flagellated chambers communicating with inhalant and exhalant canals The adult *Oscarella*
lated genus *Plakina* with two- three

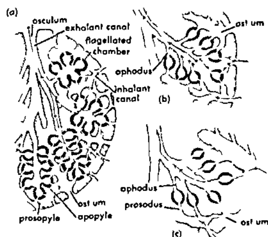


Fig 2 Types of leuconoid canal systems (a) Eurypylous chambers opening directly into exhalant canal (b) Aphodal chambers a narrow canal or aphodus leads from the chamber to the exhalant canal (c) Diplodal chambers a narrow canal or prosodus intervenes between the inhalant canal and chamber as well as between the chamber and the exhalant canal (*After Hyman 1940*)

rayed spicules retain the simple leuconoid structure The sponge consists of a folded wall each fold made up of a dermal layer and a group of flagellated chambers opening into an exhalant canal In more complicated stage as seen in *Plakortis* the membrane spreads over the outer ends of the folds and subdermal cavities are developed In many species of Tetractinomorpha an extensive cortex is developed consisting of a network of fiber cells called desmacytes which in some cases is overlaid by a thick gelatinous layer containing amoebocytes In form the species of these orders may be thinly encrusting or massive but often they have spherical or ovoid shapes Branching species rarely occur

Skeleton The skeletal system consists of spicules (sclerites) spongin fibers or both

Spicules The spicules of Demospongiae are intracellular secretions of scleroblasts cells derived

fiber Above a minimal concentration of silica in the surrounding medium the length of spicules is independent of the amount of silica present The thickness of the spicules varies up to a maximal value in correlation with silica concentration however Tetraxonid spicules are apparently also formed in individual scleroblasts Microscleres are formed in special scleroblasts they require a higher minimal content of silica for formation than do megascleres

Spongin Spongin is secreted by amoebocytes called spongioblasts which line up in rows Each cell secretes a spongin rod which fuses with adjacent rods to become a long fiber The fiber is freed following the degeneration of the spongioblasts

Ceractinomorph skeletal elements Among the ceractinomorph sponges spongin tends to be of common occurrence In the orders Dendroceratida and Dictyoceratida, the latter of which includes the commercially valuable bath sponges spicules are absent and the skeleton is formed of spongin fibers only In the orders Haplosclerida and Poecilosclerida varying quantities of spongin occur along with siliceous monaxonid spicules In some species a network of spongin fibers occurs in which the spicules are embedded, in others spongin serves as an interspicular cement In the order Halichondrida spongin is rarer in occurrence but is always present in the form of short tracts or as a cement The cement helps to unite the irregularly arranged siliceous monaxonid spicules

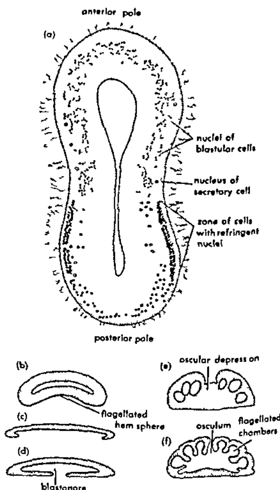


Fig 3 (a) Free swimming amphiblastula larva of *Oscarella* (b) Newly settled larva (c) Gastrulation begins by invagination of flagellated hemisphere (d) Later stage of gastrulation blastopore closing (e) Formation of osculum (f) Young leuconoid stage is formed (*After Meewis, 1939 and Lévi 1956*)

Ceractinomorph microscleres are commonly C- or S-shaped (sigmas) bow shaped (toxas) or anchor shaped (chelas) or are fine and hairlike (rhaphides)

Tetractinomorph skeletal elements Tetractinomorph sponges of the orders Homoscleropharida and Choristida have little or no spongin and almost always have some tetraxonid siliceous megascleres to which may be added monaxonid types When triaenes are present they usually occur in tracts *radiating from the central part of the sponge to the* surface

The order Clavaxinellida is characterized by a skeleton of monaxonid siliceous megascleres accompanied by varying quantities of spongin The spicules characteristically occur in tracts radiating from the central regions of the sponge or in the case of those with an abundance of spongin a plumose arrangement is found Microsclere types include asters with numerous rays diverging from a central point streptasters spiny rods often twisted spirally sigmas and rhaphides The same range of shapes occurs as in other tetractinomorph orders and in addition many species have an upright branching form

Lithistid sponges Some species of Demospongiae are characterized by the presence of spicules called desmas *These are formed by the secondary deposi*tion of silica around ordinary monaxonid or tetraxonid spicules Supplementary knobby branches

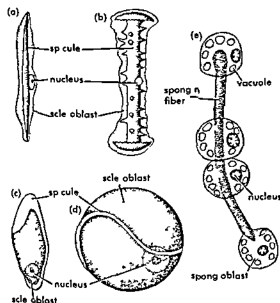


Fig 4 Skeleton formation in the Demospongiae (a) Diactinal megasclere (b) Amphidisc forming within a scleroblast (*after Evans 1901*) (c) Chela formation within scleroblasts (d) Sigma formation within scleroblasts (*from Minchin 1910 after Schmidt 1875*) (e) Spongin fiber being formed by spongoblasts in *Haliclona* (*after Tuzet 1932*)

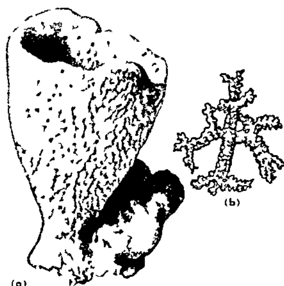


Fig 5 (a) *Discodermia ornata* a lithistid sponge (b) Desma of same (*After Sollas 1888*)

often develop and articulating processes may occur by which neighboring desmas become interlocked to form a stony or lithistid skeleton Because of the rigidity of the skeletons of lithistid sponges they are commonly preserved as fossils and are the best known Demospongiae in the fossil record Paleontologists have tended to classify such *sponges in an order Lithistida It is apparent from* studies of Recent species with a lithistid skeleton that this modification has arisen many times in the *evolution of the Demospongiae and that the order* Lithistida has no validity It is difficult to place fossil lithistids among the several orders of the class Demospongiae *however unless developmental* stages of the peculiarly modified spicules are present [W D H]

Bibliography See PORIFERA

Dendroceratida

A small order of sponges of the class Demospongiae Members of this order either have a skeleton of spongin fibers or lack a skeleton The spongin fibers when present are typically dendritic in form seldom anastomosing to form a network and arise from a basal plate of spongin adherent to the substratum The fibers which in most genera lack foreign inclusions such as sand grains are made up of concentric layers of spongin new layers apparently being added throughout the life of the sponge The flagellated chambers are large and sac shaped and open directly into the exhalant canals without the intervention of a special channel In *Halisarca* a genus without a fibrous skeleton the flagellated chambers are tubular and often branching Members of this order are mostly small encrusting sponges but some are massive or

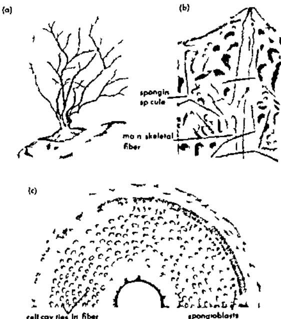


Fig 1 Skeletal features of Dendroceratida (a) Dendritic skeleton of *Dendrilla cactus* (b) Section through *Darwinella australiensis* showing concentric growth of main skeletal fibers and spongin spicules (c) Cross section of fiber of *Ianthella flabelliformis* (After Lendenfeld, 1889)

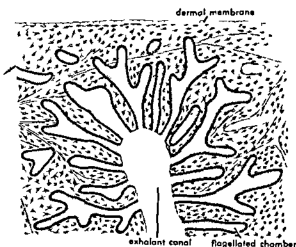


Fig 2 Section through *Halisarca dujardini* (After Lendenfeld 1889)

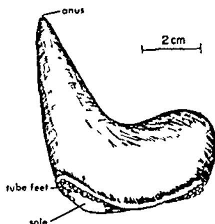
Ianthella forms laminate or vase shaped colonies Dendroceratid sponges occur chiefly in tidal and shallow coastal regions of all seas

The genus *Darwinella* possesses diactinal or

[W D H]

Dendrochirota

An order of Holothuroidea in which the tentacles assume a complex branching treelike form and are highly extensible Tube feet and respiratory trees are present and the pharynx usually has retractor muscles Plankton or detritus becomes entangled in the tentacles, which are then applied one



Psolus phantopus (After T Mortensen, 1927)

by one to the mouth and the food is sucked in Fish often attack them either swallowing them whole or browsing on the conspicuous tentacles There are two families (1) Cucumariidae, comprising numerous species of cylindrical form such as *Cucumaria*, which live in all seas at various depths among seaweeds or under pebbles, and (2) Psolidae including only a few genera of small inconspicuous species such as *Psolus*, provided with a ventral sole which serves as a creeping foot, or as a suction cup for attachment to the undersides of stones or shells *See* HOLOTHUROIDEA [H B F]

Dendrochronology

Measurement of time by trees especially by tree rings in conifers A single growth increment or growth layer (tree ring on a two dimensional cross section of a conifer) possesses two parts an earlier one formed of light colored large thin walled cells (lightwood), and a later one formed of dark colored small thick walled cells (densewood)

Basis of dating. The reading and dating of growth layers require that there be a major annual periodicity in growth factors and a recognizable response in the xylem corresponding to the periodicity Dating also requires one or the other of two assumptions either a tree must lay down only one sharply bordered growth layer per year, or the annual increment no matter how multiple or partial in areal extent must be positively identified Prolonged field studies into the growth habits of the trees of any area can eliminate one of these assumptions Some students in the southwestern part of the United States consider that accurate dating

depends further upon distinct visual variations in the thickness of growth layers in a sequence thin growth layers (some microscopic) being interspersed among thicker growth layers

Technique of dating Plotting, crossdating and chronology building are principal components of the dating technique Strikingly thin growth layers and their locations in the sequence serve as diagnostic criteria for the construction of skeleton plots

Plotting is done on suitable coordinate paper where the horizontal axis is time in years and the *vertical axis is relative thinness of certain growth* layers in arbitrary scale Figure 1 shows skeleton plots made by "reading" growth layers the height of each plotted line corresponding to the thinness of the growth layer in relation to adjacent growth layers In practice perhaps one in ten growth layers has actual diagnostic value because of consistent and conspicuous thinness



Fig 1 Skeleton plots of three tree cases crossdating and construction of a master sequence below The longer the plotted line is, the thinner the growth layer in relation to adjoining growth layers

Crossdating is the correlation or matching of different sequences (from different trees) and is chiefly dependent on the presence of thin growth layers and the number of growth layers between them Figure 2 illustrates by profile diagrams crossdating on actual wood specimens whereas Fig 1 shows crossdating by means of skeleton plots a simple convenient and effective method where many specimens are involved Not all trees in a single locality will crossdate—at some places or times only a few will do so Locally trees crossdate which have experienced the effects of similar growth factors similar fluctuations of growth factors similar genetic backgrounds and similar pathologic and physiologic histories Growth factors in

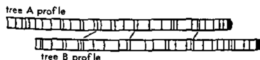


Fig 2 Crossdating on wood specimens by means of thin growth layers and the intervals between them shown on two radial profile diagrams

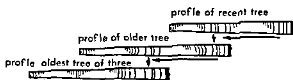


Fig 3 Diagrammatic example of chronology building by extension from one sequence to another of three radial profiles

fluencing the amount of growth in trees fluctuate within a major growing season as well as from year to year

Chronology building includes the construction of a tree ring calendar or master plot by merging the records of trees or their skeleton plots which have been crossdated and the extension of the merged record backward in time by crossdating successively older and overlapping sequences (*or skeleton plots) onto the master chart of known dates* Figure 3 illustrates extension backward in time on wood samples with successively older growth layers Beams recovered from ancient Indian dwellings in the Southwest of the United States have yielded a master chart of the Pueblo area reaching back more than 10 centuries This was done by extending the records of living trees backward step by step and by bridging the gap between a relatively dated master plot and the early end of a dated master After the construction of such a Pueblo master chart other ruins can be dated provided that they contain wood or charcoal with readable growth layers that the original trees used by the Indians grew under conditions similar to those affecting the trees which gave rise to the master chart and that the wood or charcoal belongs to the fraction of recovered materials which cross

merged record derived from structures in the same Pueblo area The validity of all such dates rests *upon four factors accurate identification of the annual increment* amount and certainty of overlap in chronology extension integrity of the master plot as a regional representative and the impossibility of an unknown sequence crossdating at more than one place on the master or standard chart

Crossdating is both interesting and most successful in the lower forest border of the United States Southwest Here within the border area growth layers are not the simple rings commonly seen on the end of a board Actually growth layers show a great variety of forms—they are highly complex Such complexity appears to be the result of more or less violent fluctuations among the growth factors especially soil moisture which are typical of the lower forest border only a short distance from semidesert conditions Growth may be impeded restricted to portions of a tree or stopped completely one or more times a year The variety of forms and

the multiple growth intervals among growth layers increase in trees situated outward from the forest interior toward the forest edge

In the extreme lower forest border of western Texas the growth layers in the trees native or introduced illustrate variety of form and multiplicity of growth intervals within a season Growth layers may be entire over the body of the tree or partial over a small or large part of the tree They may be

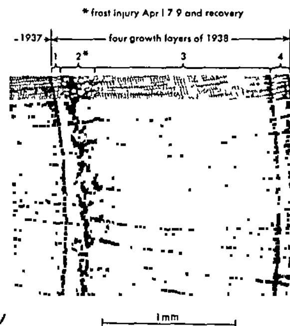


Fig 4 Transverse section of Arizona cypress wood cut December 1938 at Lubbock Texas

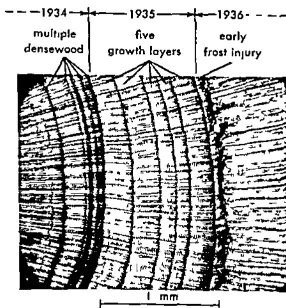


Fig 5 Transverse section of Arizona cypress wood from Lubbock, Texas

complete with both light and densewood present or incomplete with lightwood or densewood only Partial growth layers (locally absent) may be lenses half lenses arcs and their variations Outer margins of growth layers may be invisible or visible The visible margins may be diffuse, definite, or sharp A growth layer may represent an entire growing season (annual) or merely a portion of a single growing season Multiplicity of growth layers in the annual increment modified by the genetic history of the tree species, characterizes tree growth in the lower forest border Figures 4 and 5 illustrate this multiplicity in entire growth layers Among lenticular growth layers only three cases out of several thousand are found to be annual

Such multiplicity controlled by genetic differences and by position of the original trees in relation to the forest border, modifies but does not destroy the dating in the Pueblo area of the Southwest The number of intra annual growth layers, which were assumed to be annual when the beams from Indian dwellings were dated, is estimated to reach a maximum of 15 per century and to average about 5% If a chronology has accuracy and validity any dates derived by its use are in spite of inherent multiplicity remarkably close to the truth [w s c.]

Bibliography A E Douglass, The secret of the southwest solved by talkative tree rings *Natl Geog Mag*, 56 736–770, 1929, W S Glock, Cambial frost injuries and multiple growth layers at Lubbock Texas *Ecology*, 32 28–36 1951, W S Glock *Principles and Methods of Tree ring Analysis*, Carnegie Inst Wash Publ 486 1937

Dendrology

The division of forestry concerned with taxonomy of trees and other woody plants Dendrology, called forest botany in some countries usually is limited to taxonomy of trees but may also include shrubs and woody vines This basic subject in the training of foresters teaches how trees are named (nomenclature) described (morphology), and grouped (classification) how to find the name of an unknown tree and recognize important forest species (identification), and where trees occur both by geographic ranges of species and by forest types (distribution)

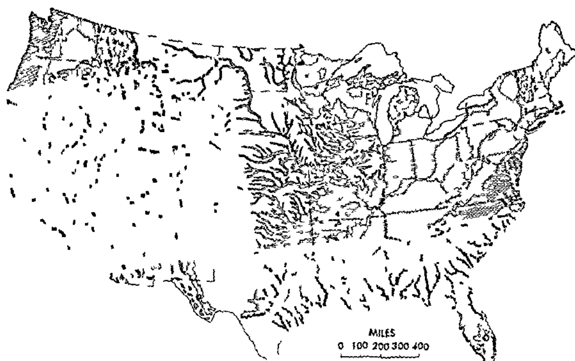
In forestry a tree is defined as a woody plant which has a single erect perennial stem or trunk at least 3 in in diameter at breast height (4½ ft above ground) is not less than 12 ft tall, and supports a definitely formed crown of foliage A shrub is a woody plant generally lower growing than a tree and frequently having several slender perennial stems arising at or near the ground

Common names of trees and lumber in the United States have been standardized by various forest agencies The scientific name of a tree as in other plants, is in Latin and consists of two parts, a genus and a species, for example, *Pinus ponderosa* ponderosa pine

Trees are described and distinguished in botan ical terminology largely upon their characteristics of *form* and structure (*morphology*) Principal parts useful in identification are leaves flowers fruits seeds buds twigs branches trunk bark and wood (*see* PLANT CLASSIFICATION) (See sepa rate articles on trees listed under common names)

Classification of forest trees A common but artificial classification groups plants into trees shrubs, and herbs However the 50 000 or more tree species in the world as well as the numerous other species making up the plant kingdom are ar ranged scientifically according to natural relation ships as indicated by evolutionary evidence (*see* PLANT KINGDOM) For example several genera having a number of characteristics in common compose the pine family These and trees of related families make up the gymnosperms sometimes re ferred to as softwoods by lumbermen (*see* PLANT ANATOMY) Most tree species however are flower ing plants (angiosperms) and nearly all of these are dicotyledons often called hardwoods Palms and bamboos are monocotyledons

The approximately 785 species of trees in the United States are grouped into 221 genera and 69 plant families However only about 175 species in 50 genera are commercially important for lumber (*see* LUMBER MANUFACTURE) The larger total cov



FOREST VEGETATION (WESTERN)

- Spruce F r (northern con ferous fo est)
- Cedar Hemlock (northwestern con ferous fo est)
- Western larch–Western white p ne
  - Fac fc Douglas f r
  - Redwood
- P nyon-Jun per (sou hwestern con fe ous woodland)
- Chapa al (southwes e n broad eaved woodland)
- Ponderosa p ne–Douglas f r (western p ne fo est)
  - Ponderosa pine–Sugar p ne
  - Ponderosa p ne–Douglas f r
  - Lodgepole p ne

FOREST VEGETATION (EASTERN)

- Spruce Fir (northern con ferous fo est)
- Jack Red and Wh te P nes (northeastern pine forest)
- Bi ch Beech-Maple Hemlock (northe n ha dwoods)
- Oak (southern ha dwood forest)
- Chestnut–Chestnut Oak–Yel ow poplar
  - Oak Hickory
  - Oak P ne
- Cyp ess-Tupelo Sweetgum (r ver bottom forest)
- Longleaf Loblolly and Slash p nes (southeaste n p ne fo est)
- Mang ove (subt op cal fo est)

Areas character zed by major forest types in the Un ted States (*U S Forest Serv ce*)

ers several small trees including 148 mostly shrubby species of hawthorn (*Crataegus*) more than 100 subtropical and tropical species of southern Florida and others of limited occurrence or low quality wood Important forest trees are described in separate articles listed under their common names

Identification of forest trees The correct scientific name of a recorded species may be determined by means of printed keys or manuals or by comparison with a known tree or with mounted specimens in a herbarium (*see* PLANT KEYS) Even in winter leafless trees usually can be identified from keys by studying their bud twig and bark characteristics Nearly every state publishes a popular inexpensive illustrated pocket guide or bulletin for the identification of the trees of that state Regional floras are covered in other publications

Distribution of forest trees Each tree species has its own natural distribution or range A few tree species of the United States extend across the continent while others are local and rare in distribution About two-thirds of the important forest trees are eastern or southeastern whereas one third are western some in the Rocky Mountains and others in the Pacific Coast region Trees also have an altitudinal distribution and zonation in high mountains *See* FOREST ECOLOGY

Forest stands of similar composition appearance and structure are grouped together into areas characterized by major forest types or formation and are named from the predominant or characteristic species The Society of American Foresters has defined 106 forest cover types of eastern North America (exclusive of Mexico) and 50 of western North America *See* FOREST CONSERVATION PLANT TAXONOMY [E L L]

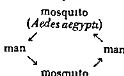
Bibliography *See* FOREST AND FORESTRY

Dengue fever

An acute viral disease with fever rash prostration and lymphadenopathy Inapparent infections are frequent complications or deaths are rare

The virus is a member of arbor group B At least two antigenic types exist It is difficult to isolate intracerebral inoculation of infant mice is the only generally satisfactory method for producing consistent measurable viral damage *See* ANTIGEN ARBOR VIRAL ENCEPHALITIDES

The human infection cycle is



Monkeys and *A albopictus* possibly form the jungle reservoir Outbreaks occur chiefly in Africa India the Far East and also in Hawaii the Philippine and Caribbean islands and the eastern Mediterranean area Destruction of *A aegypti* in the American tropics may be eliminating dengue *See* ANIMAL VIRUS [J L M]

Density

The density of a given material is defined as the mass per unit volume of the material The term is applicable to mixtures and pure substances and to matter in the solid liquid or gaseous state Density of all matter depends on temperature the density of a mixture may depend on its composition and the density of a gas depends on its pressure Common units of density are grams/cm³ and slugs or pounds per ft³ The specific gravity of a material is defined as the ratio of its density to the density of some standard material such as water at a specified temperature for example 60°F or for gases air at standard temperature and pressure Another related concept is weight density which is defined as the weight of a unit volume of the material *See* DENSITY MEASUREMENT MASS, WEIGHT [L N]

Density measurement

The measurement of the mass per unit volume of a material Density is usually expressed in grams per cubic centimeter (or milliliter) pounds per cubic foot or pounds per gallon The densities of materials change with temperature and pressure and sometimes with other ambient conditions such as humidity so that conditions of measurement should be specified Weighing is normally done in air with a density of about 0 0012 g/ml thus precision density determinations require a correction for this buoyant effect The correction is usually insignificant except in gas density determinations

The specific gravity of a liquid or solid is the ratio of the density of the substance to that of water at a specified temperature In scientific work the reference is usually water at 4°C in engineering work it is usually water at 60°F The specific gravity of a gas is the ratio of its density to that of dry air (usually at 0°C and 760 mm Hg) In the metric system density and specific gravity with a 4°C water reference have the same numerical value and differ only by 0 1% with a 60°F reference In the English system the numerical values of density and specific gravity for the same material are quite different

In industry and in the laboratory specific gravity determinations are more common than density measurements Furthermore the measurement of liquid gravities dominates that of gases and solids because this characteristic is convenient in determining the quality strength or composition of many liquids Therefore the hydrometer which is applicable to liquid specific gravity measurements exclusively is the most widely used specific gravity instrument (*see* HYDROMETER) The other common density determining methods are outlined below

Solids and gases It is inconvenient and seldom desirable to measure the density of solids and gases while they are being processed in industrial plants Density determinations of gases and solids are made in the laboratory on a fixed sample of the material An accuracy to one part in 1000 is easily

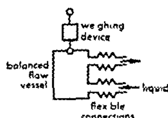


Fig 1 Balanced flow vessel for continuous density measurement (D M Considine, ed, Process Instruments and Controls Handbook, McGraw-Hill, 1957)

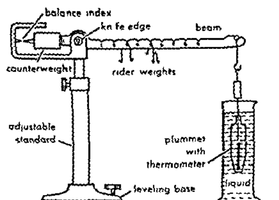


Fig 2 Westphal balance (R J Sweeney Measurement Techniques in Mechanical Engineering Wiley 1953)

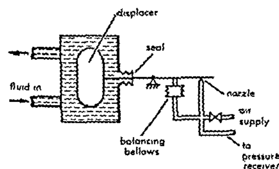


Fig 3 Buoyancy type density transmitter (D M Considine, ed, Process Instruments and Controls Handbook, McGraw Hill, 1957)

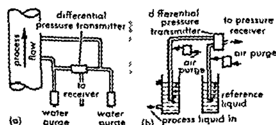


Fig 4 Specific-gravity measurement (a) With water purge (b) With bubbling system (From D M Considine, ed, Process Instruments and Controls Handbook, McGraw Hill 1957)

obtained and better accuracy is possible with precision measurement and the application of known corrections

Liquids. Liquid density is also measured by weighing a known volume (pycnometer method) By using flexible connections to the volume, the same principle can be used for a continuous indication of liquid density (Fig 1) The continuous instrument is sensitive to density changes of less than 0 001 g ml and its time constant can be as short as 1 min depending upon the sampling system

The Westphal balance (Fig 2) measures liquid density by the difference in weights of a solid of known volume in air and immersed in the liquid (buoyancy method) It is also used to determine the specific gravity of solids by immersion in a liquid of known specific gravity A precision balance provides accuracies to one part in 10 000 This same principle (Fig 3) is used industrially for a continuous indication or record of liquid density A sensitivity of 0 001 g/ml can be achieved with careful design and a slow sample flow rate

The differential pressure manometer may be used with water purges (Fig 4*a*) or with an air bubbling system (Fig 4*b*) to measure density With these or similar systems it is difficult to achieve a dependable sensitivity of even 0 01 g/ml on a carefully engineered and maintained installation

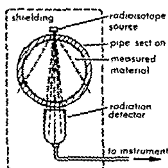


Fig 5 Density by γ-ray measurement (Industrial Nucleonics Corp)

Density of liquids may also be measured by the absorption of γ rays within the liquid This method, while expensive, is particularly advantageous when the liquid is flowing in a line or standing in a tank which should not be disturbed The γ-source is placed on one side of the line or vessel and the detector on the opposite (Fig 5) Both must be well shielded, and a sensitive and stable electronic amplifier is required to detect small density changes See GAMMA RAYS [R E CL]

Dentistry

That part of medical science which is concerned with the prevention, diagnosis, and treatment of diseases of the teeth and adjacent tissues and the restoration of missing dental structures It is a division of a total health service which requires a

thorough knowledge of the structure origin growth and function of the oral cavity and its related structures Also it requires an understanding of the relationship of the oral cavity to other parts of the body in health and disease and a recognition of the manifestations of diseases of the body which are reflected in the mouth Although there are seven branches of specialization in dentistry the greatest portion of the dentist s time is spent in restorative dentistry

Restorative dentistry Restorative dentistry is that phase of dentistry which aims to conserve and restore the defective tooth The dentist practitioner of dentistry directs the greatest proportion of his time and efforts towards this phase of dentistry because dental caries (tooth decay) is one of the most prevalent diseases known to medical science This important division of dentistry accounts for the autonomous position the profession enjoys as a part of a total health service The dentist must possess a thorough knowledge of basic and medical sciences he must learn some principles of engineering and metallurgy and he must acquire an unusual degree of manual dexterity and digital skill in the performance of restorative dentistry

Some causes of tooth destruction are congenital defects erosion abrasion traumatic injuries and caries One must also consider the improvement of function esthetics and speech impairments in relation to the teeth

In recent years there has been an accelerated improvement in dental instruments and materials which has benefited both the patient and the dentist By the use of carbide burs and diamond stones and points combined with the ultrahigh speed revolving instruments tooth structure may be removed with a minimum of discomfort to the patient Also ultrasonic vibrations are now being used to remove foreign material and stains from the teeth thus reducing tedious hand instrumentation to a minimum In addition new restorative and impression materials have increased the efficiency accuracy and performance of dental restorative services

Specialization Specialization in the practice of dentistry has developed as a result of the increasing complexity of dental science Oral surgery orthodontia pedodontia periodontia prosthodontia oral pathology and public health dentistry are the branches of dentistry recognized by the American Dental Association

Oral surgery This is the branch of dental science concerned with the treatment of diseases growths and injuries involving the oral cavity and related structures by means of manual operations and therapeutic measures Some of the cases treated by the oral surgeon are jaw fractures teeth requiring extraction neoplasms cysts and congenital defects Modern instruments drugs and x rays have facilitated this phase of dentistry

Orthodontia Orthodontia deals with the prevention and correction of irregularities of the teeth and jaws and the resultant facial disharmony The objective is to produce normal occlusion of the teeth and facial harmony Relocation of teeth produced by mechanical appliances and elastic traction or by correction of abnormal habits is maintained for a period of time to allow for bone growth and muscle changes The final result depends upon correct positioning of the teeth normal function of the muscles bone production and general health A normal relationship of the teeth and jaws helps in giving shape and expression to face and mouth aids in enunciation and sounding of words and permits the proper mastication (chewing) of food

Pedodontia This branch of dentistry is concerned with the dental care and treatment of children It cannot be denied that effective dental prevention must start with the child and that once major destruction of the teeth gums and alveoli is in evidence it is too late for this method

The necessity of maintaining the deciduous (baby) teeth is evident these teeth are an essential part of nature s plan of development The deciduous teeth were so placed to accommodate the small arches and face and to allow for proper chewing of foods As the child continues to grow vertically and laterally the permanent and larger teeth replace the deciduous teeth

Periodontia Periodontia is that branch of dentistry devoted to the study prevention and treatment of diseases which affect the supporting tissues [illegible] (commonly [illegible]

irritation of the gingiva (gums) about the necks of the teeth with mechanical irritants such as calculus (tartar) on the teeth food impaction through improper positioning of teeth poor dental restorations traumatic occlusion and chemical irritants This form which begins at the margin of the gingivia is called gingivitis it gradually involves the deeper periodontal tissues until all the supporting tissues of the tooth are involved Through destruction of the periodontal tissues and bacterial invasion the tooth is loosened and usually lost Treatment depends upon the stage of the disease removal of all irritants and the general health of the individual or extraction of the tooth or teeth involved Preventive treatment through strict oral hygiene and dental prophylaxis has proved to be the most effective way of controlling this disease The second type of periodontitis consists of a gradual destruction of the supporting tissues of the teeth without apparent cause It has been attributed to unknown systemic conditions

Prosthodontia This is the branch of dentistry devoted to the construction and replacement of oral structures with an artificial substitute The replacement of teeth is necessitated by congenitally missing teeth or loss of teeth by extraction accident or disease Dentures are either partial or complete the former being used when some teeth are present and the latter when all the teeth are missing Bridges are used when there are teeth present at

each end of the missing space to support the prosthesis and may be fixed or removable depending upon the position, strength of supporting teeth and other pertinent factors The dentist constructs special prosthetic appliances to replace oral structures missing as a result of surgery, accidents, and congenital defects

Oral pathology Oral pathology is concerned with the diagnosis and treatment of diseases of the teeth, jaws, and oral mucosa Also it recognizes the manifestations of other diseases throughout the body which are reflected in the oral cavity An oral pathologist must correlate the etiological histological roentgenological, and clinical picture into a pattern of disease from its inception through its termination

The pathology of the oral cavity is divided into the following six etiological classifications (1) the infectious diseases, including that pathology manifested by a bacterial fungus or viral invasion of the tissues, (2) those conditions present as the result of noninfectious diseases such as inflammatory, dystrophic or congenital disturbances, (3) the metabolic and endocrine disturbances throughout the body, (4) the nutritional and dietary disturbances which are reflected in the oral cavity, (5) those conditions caused by chemical or physical injuries, and (6) the pathology of the oral cavity when there are disturbances of the blood forming organs

Dental public health This aspect of dentistry is defined as the science and art of preventing and controlling dental diseases and promoting dental health through organized community efforts It is the sum total of research education prevention diagnosis, prescription treatment and evaluation in community dental care Public health dentists must have skill in public and human relations a knowledge of administrative practice and the ability to determine the communities' needs for diagnostic and dental care facilities so as to assist in their design development, administration and support They must have a knowledge of the principles methods and media available in dental health education and must advise and assist in the development of dental health educational programs They must possess a broad understanding of the social and behavioral sciences as they affect the promotion of health and prevention of disease in the community This group has sponsored in many communities the fluoridation of water systems which has been an aid in controlling caries

Research A new era of dental research came into existence on June 14 1948 when both houses of the Eightieth Congress of the United States passed a bill providing for the establishment of a National Dental Research Institute as part of the National Institutes of Health With the establishment of the National Institute of Dental Research along with the already existing research center of the National Bureau of Standards private research groups and research educational foundations of leading dental schools there are indications at last dental research is geared to deal with the problems of dental caries and periodontal disease

Fundamental studies are now underway in such fields as oral bacteriology human genetics pathology, biochemistry, epidemiology, and biophysics Reports emanating from these various studies suggest a greater understanding of the so-called elusive etiology of such disorders as caries and periodontal disease *See* BACTERIOLOGY, MEDICAL, BIOCHEMISTRY, BIOPHYSICS DISEASE, EPIDEMIOLOGY, HUMAN GENETICS, MEDICINE, PATHOLOGY, TOOTH [I L H , V H M]

Bibliography J C Brauer et al, *Dentistry for Children*, 4th ed 1958, *Bulletin of the American Board of Dental Public Health*, 1951, L W Burket, *Oral Medicine*, 2d ed, 1952, W B Dunning and S E Davenport Jr., *A Dictionary of Dental Science and Art*, 1936, L I Grossman (ed), *Lippincott's Handbook of Dental Practice*, 3d ed , 1959, Michigan University School of Public Health Objectives and evaluation of a state's dental program *Proc 4th Workshop on Dental Public Health* ser 67 1956 B J Orban and F M Wentz *Atlas of Clinical Pathology of the Oral Mucous Membrane* 2d ed 1959

Dentition

In animals the arrangement type, and number of teeth which are variously located in the oral or pharyngeal cavities Teeth are found in areas where there is an underlying supporting structure of cartilage or bone or where stomodeal ectoderm is present The bones with which these structures are usually associated are the mandible premaxillaries and maxillaries However in certain species the vomerine palatine parasphenoid and pterygoid bones may be involved

Attachment of teeth is variable among vertebrates They may be inserted in sockets in the jaw bones (thecodont condition), fused to the upper edge of the bone proper (acrodont condition) or attached to the inner surface of the jaw bone (pleurodont condition) In polyphyodont animals, teeth may be constantly replaced Most mammals have two sets of teeth called a diphyodont condition during their lifetime These are the deciduous or milk teeth and the permanent dentition Monophyodont dentition is the development of only one set of teeth

Teeth may have a similar form in all regions where they occur Such teeth are homodont as distinct from those that are variable in shape, called heterodont dentition The heterodont condition is frequently described by a dental formula expressing both the number and kind of teeth in each half jaw both upper and lower For man the following pertains

$$i\frac{2}{2}\quad c\frac{1}{1}\quad p\frac{2}{2}\quad m\frac{3}{3}$$

This can be read that for either the right or left side of the jaw in man, there are 2 upper lower incisors 1 upper and lower canine,

Dental formulas of some vertebrates

Animal	Teeth i	c	p	m	Total
Man	2	1	2	3	32
	2	1	2	3	
Cony	3	1	4	4	48
	3	1	4	4	
Beaver	1	0	1	3	20
	1	0	1	3	
Cat	3	1	3	1	30
	3	1	2	1	
Dog	3	1	4	2	42
	3	1	4	3	
Sheep	0	0	3	3	32
	3	1	3	3	
Lynx	3	1	2	1	28
	3	1	2	1	
Rat	1	0	0	3	16
	1	0	0	3	
Horse	3	1	4	3	44
	3	1	4	3	
Mole	3	1	4	3	44
	3	1	4	3	
Squirrel	1	0	2	3	22
	1	0	1	3	
Reindeer	0	0	3	3	32
	3	1	3	3	
Pig	3	1	4	3	44
	3	1	4	3	
Common seal	3	1	4	1	34
	2	1	4	1	
Skunk	3	1	3	1	34
	3	1	3	2	
Raccoon	3	1	4	2	40
	3	1	4	2	
Bear	3	1	4	2	42
	3	1	4	3	

molars and 3 molars According to this formula man has a total of 32 teeth located in the following manner 8 teeth in each half of the upper jaw and the same number in each half of the lower jaw Usually the initial letter is omitted in the formulas and if a tooth does not appear in the group a zero indicates this fact *See* TOOTH [C B C]

Deoxyribonucleic acid

One of the two nucleic acids also referred to as DNA A part of the DNA molecule is a sugar D-2 deoxyribose The other nucleic acid is ribonucleic acid (RNA) whose molecule contains the sugar D-ribose DNA is the primary chemical carrier of heredity This article discusses the genetic role of DNA For information on the chemistry of DNA *see* NUCLEIC ACID

In contrast to RNA DNA is metabolically inert once synthesized it is not easily broken down It is essential for cell division Interference with DNA synthesis stops cell division even though the synthesis of RNA and of protein may go on *See* RIBONUCLEIC ACID

A genetic role for DNA was indicated by its localization in the chromosomes its high concentration in sperm cells (spermatozoa contain no RNA) and the fact that certain mutagenic agents such as nitrogen mustards x rays and ultraviolet radiation were shown to alter chemically the DNA and no other cellular constituent *See* CHROMOSOME MUTATION

The most direct proofs however, that DNA per se is capable of transmitting genetic information come from studies with bacteria The bacterial viruses or phages (bacteriophages) for example attach themselves to the bacterial surface but inject only their DNA and not their protein coat into the bacterium This DNA can then bring about the production of many new phage particles Still better evidence comes from the phenomena of bacterial transformation and transduction In both DNA is transferred from a donor to a recipient cell and can permanently confer on the latter some property of the donor The characters most commonly transferred are the ability to form a capsule drug resistance and the ability to form certain enzymes In transformation the transfer occurs in vitro by means of isolated chemically purified DNA In transduction the transfer occurs in vivo by means of a virus that is nonlethal and serves as a carrier of the genetic material from a donor cell to a recipient cell There is also a bacterial mating process in which the donor injects into the recipient cell a piece of its chromosome which as has been shown by tracer studies is DNA *See* BACTERIOPHAGE

The exact manner in which DNA stores and transmits genetic information is still unknown However there is good evidence that it is stored linearly along the chromosome When bacterial mating for example is interrupted after a short piece of chromosome has been injected only a small proportion of the characters contained in the whole chromosome is transferred It is thought that this information is coded in specific arrangements of the four different nucleotides found in DNA much as the letters of the alphabet can be arranged into thousands of words each with a particular meaning and which in turn can be arranged into sentences So far the 'code of DNA has resisted all attempts to crack it but there is evidence that a sequence of some 3 to 5 nucleotide pairs corresponds to a particular amino acid in the structure of a specific protein *See* NUCLEOPROTEIN

In considering how DNA transmits its information there are two main problems First it must be able to replicate itself exactly second it must control the synthesis of some cytoplasmic agent which can translate the information into enzymes which in turn control all other metabolic activities of the cell

A widely accepted working hypothesis for the mode of DNA replication is the following The DNA molecule consists of two long strands of deoxypolynucleotide wound around each other in a helical fashion and held together by means of hydrogen bonds The two strands are complementary in that the positions for the bases adenine and thymine in one chain correspond to guanine and cytosine in the other and vice versa At the time of replication the two chains separate and each directs

the synthesis of its complement thus giving rise to two complete molecules This hypothesis is purely formalistic in that it does not tell how the actual synthesis of the new strands occurs and although there is evidence that in every generation one-half of the DNA is transferred intact to the daughter cell and may serve as a template on which a complementary strand of nucleotides is laid down the hypothesis may be too simple to account for all known facts [I D R]

Depolymerization

The reversion of a polymer to a monomer as shown below for a trimer of propylene 2 4-dimethyl-4-heptene.

$$CH_3-\underset{|}{\overset{CH_3}{CH}}-CH_2-\overset{CH_3}{\underset{}{C}}=CH-CH_2-CH_3 \rightarrow 3CH_2-CH=CH_2$$

Depolymerization normally is not employed as a distinct process in the petroleum industry but may occur incidentally during catalytic or thermal cracking operations if the feedstocks contain polymers

The reaction is accelerated by protonic catalysts notably inorganic acids placed on porous supports or acidic cracking catalysts such as silica alumina at temperatures of 250–500°C and is then identical in mechanism with the catalytic cracking of olefins Depolymerization by heat alone will proceed by a thermal cracking mechanism In both cases the product may contain substantial quantities of polymer fragments differing from the original monomer *See* CRACKING POLYMERIZATION [B S G]

Dermaptera

An order of insects which are elongate 60 mm to 190 mm in length with incomplete metamorphosis and chewing mouth parts They are known as the earwigs The European earwig *Forficula auricularia* L a species introduced into the United States has proved to be a garden pest and a household nuisance

Earwigs are nocturnal terrestrial animals which live in moist shady places under stones boards bark of trees in cracks in the soil piles of refuse hollow stems or in any other secluded and protected place

They have tough shiny bodies and slender antennae The cerci at the tip of the abdomen are modified into a pair of strong movable forceps Earwigs are apterous or winged When present the forewings are short and leathery and at rest conceal all but the tips of the folded hindwings The latter are large semicircular and membranous and have radial veins The order contains about 1100 species

In cooler regions they appear to be mainly herbivorous living for the most part on dead or decaying vegetable matter but frequently eating parts of or entire living plants In the tropics

The European earwig ventral aspect (*From E O Essig College Entomology Macmillan 1942*)

they are often carnivorous and devour many kinds of insect larvae especially maggots

Earwigs are among the more ancient insects and date from the oldest division of the European Jurassic *See* EXOPTERYGOTA INSECTA [E O E]

Dermatitis

Any inflammation of the skin Few subjects have received as much medical and lay attention in matters of diagnosis classification treatment and causative agents as dermatitis However there is often little agreement among experts on many facets of these problems

Contact dermatitis is an acute or chronic condition resulting from irritation by or sensitization to some substance in the environment Poison ivy cosmetics foods chemicals drugs fabrics and an almost unlimited list of substances may either cause direct irritation or create a hypersensitive state in a susceptible individual Itching is a prominent symptom and is usually localized to the area of some type of eruption acute or chronic which may be characteristic of a certain agent or be nonspecific in its appearance Unfortunately remo the offending substance may not ensure of the dermatitis which in many cases

sist and spread in altered form *See* HYPERSENSITIVITY

Neurodermatitis may be localized or widespread chronic inflammation characterized by an intense itching and resultant thickening or lichenification of the skin Often a history of family susceptibility to allergic conditions is present and in many individuals there is a preceding or concomitant record of nervous tension or emotional stress

Seborrheic dermatitis is a common disease of the scalp and other skin regions in which there is a tendency for greasy scaling and variable degrees of inflammation Many factors are believed to be important in the etiology of the disease such as oily skin nutritional or hormonal imbalances and certain emotional disturbances No one of these however can be isolated as the only or true cause in most cases

Stasis dermatitis is a chronic disorder apparently resulting from poor circulation and related conditions It is found principally in older persons especially those who have an occupation which requires long periods of standing A certain familial tendency is sometimes present and the incidence of stasis dermatitis is higher among people who suffer from varicose veins

Various forms of eczema represent scaling inflammations of noninfectious but often unknown cause The most common categories are infantile occupational and senile eczemas

Many other kinds of dermatitis are recognized and may result from such diverse agents as heat chafing infections and unknown causes as in psoriasis *See* PSORIASIS [E C S T]

Dermatophytosis

An infection of the skin of man and animals caused by fungi which live in the keratinized tissues but which are unable to invade the subcutaneous or deeper tissues Classification of the infection is based upon its clinical appearance rather than upon the etiological agents because several fungi may give rise to the same type of clinical lesion For example ringworm of the scalp may be caused by any one of the species of *Microsporum* or *Trichophyton* Thus a fungus infection of the scalp is tinea capitis of the feet tinea pedis or of the glabrous skin tinea corporis

Fungi causing dermatophytoses are referred to as dermatophytes and fall into three genera *Microsporon Epidermophyton* and *Trichophyton* In the United States tinea capitis is seen almost entirely in children Approximately 90% of these infections are caused by *Microsporum audouini* the remaining 10% are caused by *Microsporum canis Trichophyton tonsurans* and other species of *Trichophyton* *See* MONILIALES

Tinea pedis is usually seen in adults rarely in children The two dermatophytes most frequently seen as the etiological agents are *Trichophyton rubrum* and *Trichophyton mentagrophytes*

Some instances of dermatophytosis may be considered an occupational disease among farmers and veterinarians Cattle and horses often have ringworm infections which go unnoticed but when an individual comes into contact with these lesions he may in turn acquire the infection

Trichophyton verrucosum and *Trichophyton mentagrophytes* are two of the dermatophytes most often seen in these instances

The epidemic type of ringworm of the scalp is that caused by *Microsporum audouini* It is passed from child to child by such methods as exchanging hats using borrowed combs or by leaning the head against the back of a chair upon which infected hairs rest However when this dermatophytosis is caused by *Microsporum canis* or *Trichophyton mentagrophytes* it is usually contracted from an infected pet or farm animal

Tinea pedis is not so contagious as many would believe It has been demonstrated repeatedly that maintenance of proper foot hygiene that is drying between the toes following a bath powdering the feet in warm weather and wearing shoes that fit properly reduces to a minimum the chance of an individual's acquiring this infection

Little is actually known about the immunology of the dermatophytoses It is not known why tinea capitis clears spontaneously as a child passes through puberty nor is there a satisfactory explanation for the rarity of tinea pedis in children Some patients will develop an allergy to the fungus which causes their dermatophytosis this is manifested as an eruption secondary to the primary infection The fingers and hands are the most common sites for these dermatophytids (ids) although they may occur elsewhere on the body This allergic reaction will clear up as the primary lesions heal *See* HYPERSENSITIVITY MYCOLOGY MEDICAL [L D H]

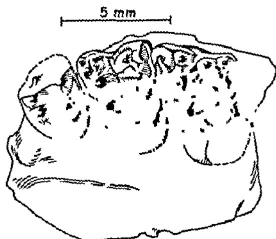
Dermoptera

An order of mammals of which only the peculiar flying lemur (*Galeopithecus* sometimes called *Cynocephalus*) of southeast Asia the East Indies and the Philippines is extant A few fragmentary fossils from the early Tertiary of North America have been provisionally placed in this order The flying lemur is apparently a very aberrant derivative of the insectivores but is so distinctive that all students agree in placing it in a separate order An extensive parachute membrane begins at the neck includes the legs to the tips of the digits and extends to the tip of the tail The animal is an expert glider but cannot fly The dentition is remarkable but is basically similar to that of insectivores The diet consists exclusively of leaves and fruits The flying lemur is the size of a large squirrel *See* EUTHERIA MAMMALIA

[D D D]

Dermoptera fossils

The Colugos or flying lemurs living dermopterans are confined to southeastern Asia Two North American genera *Planetetherium* from the late Paleocene and *Plagiomene* from the early Eocene



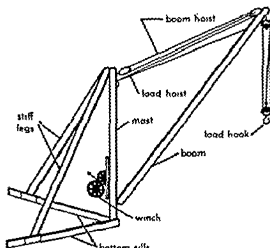
Right upper molars of *Plagiomene*, an Eocene dermopteran (After W Matthew)

are referable to the order but no other fossil dermopterans are known Fossil dermopterans are recognized on the basis of their comblike incisors and characteristically cuspidate cheek teeth

Dermopterans are believed to be descended from some unknown early member of the menotyphlan branch of the Insectivora This conclusion is supported by the early Cenozoic fossil representatives but is actually based mainly on the cranial anatomy of recent forms The earliest dermopteran, *Planetetherium*, was restricted to a swamp habitat, where it was locally abundant *See* INSECTIVORA FOSSILS. [M C MC]

Derrick

A hoisting machine consisting usually of a vertical mast, a slanted boom, and associated tackle (as illustrated) A derrick may be in any of a wide variety of forms The mast may be no more than a base for the boom, it may be a tripod, an A frame, a fixed column, and so on Fixed stays may guy it in place The boom may be fixed it may pivot at the base of the mast it may swing horizontally from near the top of the mast or it may be omitted The derrick may be permanently fixed a temporary structure or mobile on a cart or truck



Boom derrick with swinging mast and anchored stiff legs

Derricks are widely used in construction, in cargo handling, and in shops Their lifting action is intermittent compared to bucket conveyors, and their coverage is limited by the reach of the boom A manual or powered winch provides the lifting action by coiling in the running tackle the load swinging free as it rises *See* BULK HANDLING MACHINES, CRANE HOIST, HOISTING MACHINES, OIL AND GAS WELL DRILLING [D O H]

Bibliography D O Haynes *Materials Handling Equipment*, 1957

Derris

A genus of tropical shrubs belonging to the legume family (Leguminosae) These plants with their long branches climbing over other vegetation, occur as members of the jungle undergrowth in Malaysia Extracts of the roots of *Derris elliptica* have long been used by the natives as an arrow poison and to stupefy fish so they can be caught more easily Derris root is an excellent insecticide, being harmful to both chewing and sucking insects but not poisonous to human beings The insecticidal ingredient of derris root is a white crystalline substance, which is called rotenone *See* ROSALES [P D S]

Descriptive geometry

A mathematical graphical procedure that has for its purpose the visualization of structures and their exact representation in drawings After analysis of the structure each element is shown in the drawing in its exact geometrical relation to the other elements

There are two basic methods of descriptive geometry the projection method and the direct method The two methods differ as regards the attitude of mind toward the structure and toward the drawing that represents the structure

Projection method. Gaspard Monge (1756–1818) a French mathematician originated the projection method of descriptive geometry While working as a designer for the French government he was given the job of making plans for a proposed fortress This was a tedious process and involved long calculations He invented graphical solutions and completed the plans in such a short time that at first the commandant refused to accept them For a long time the graphical process was kept a state secret, finally being revealed about 1795

In the projection method, the horizontal projection plane *H* and the vertical projection plane *V* intersect in the line *GL*, called the ground line (Fig 1) These two projection planes divide into four quadrants, or angles, as num' . Illustration

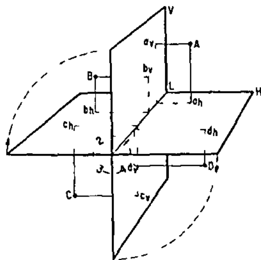


Fig 1 The planes of projection (From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw-Hill 1959)

Point A in the first quadrant is projected onto the horizontal plane at a_h by means of a projection line perpendicular to the H plane and onto the vertical plane at a_v by means of a projection line perpendicular to the V plane The projections of the points B C and D in the other quadrants are located in a similar manner If desired a profile projection plane perpendicular to the H and V planes may be introduced Right angle projection as described above is called orthographic projection

To represent horizontal and vertical projections on a flat sheet of paper the planes are conceived as

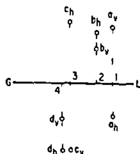


Fig 2 Projection of points (From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw-Hill 1959)

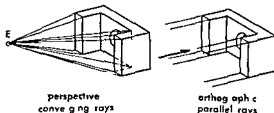


Fig 3 Two methods of viewing an object (From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw Hill 1959)

being hinged along the ground line and brought together by closing the second and fourth quadrants Projections of A B C and D then appear in a single plane (Fig 2) The H and V projections of a point are always in the same perpendicular to the ground line The usual custom in the United States is to draw objects as if they had been projected from a position in the third quadrant this is called third angle projection First angle projection is standard in some countries and professions

There are two general types of views perspective and orthographic (Fig 3) A perspective view of an object is observed from a fixed station point or point of view by means of converging rays of light that meet at the eye of the observer An orthographic view of an object is observed in a chosen direction by means of parallel rays of light

Direct method By the direct method the attention is focused on the visualized structure or object Each view of the object is obtained by looking at the object in a definite direction The view is orthographic A view never is considered as two-dimensional or as projected or drawn on a plane this is understood by the engineer who makes and reads the drawings

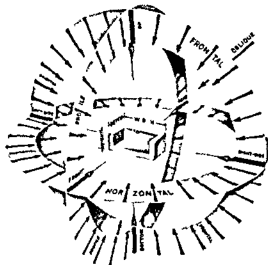


Fig 4 Viewing an object (From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw Hill 1959)

Orthographic views may be classified into three types principal views auxiliary views and oblique views The object can be viewed from any direction around three rings—horizontal frontal and profile (Fig 4) The rings represent three mutually perpendicular planes The intersections of the rings define three mutually perpendicular directions from which six principal views are observed front and rear top and bottom right and left sides

An auxiliary view can be observed around any ring in a direction perpendicular to one and only one of the directions in which principal views are

observed Auxiliary views observed in a horizontal direction are called horizontal auxiliaries or auxiliary elevations, those observed in a frontal direction are called frontal auxiliaries, and those observed in a profile direction are called profile auxiliaries

All views other than principal or auxiliary views are oblique In Fig 4 a single arrow marked oblique, indicates one of the infinite number of directions in which oblique views are observed

Dimensions of structures All structures occupy space and have three dimensions width height and depth These are measured in three mutually perpendicular directions (Fig 4)

Grouping of views In Fig 5 are shown four views of the object pictured in Fig 4 Each view is placed in relation to its adjacent view in the position from which it was viewed Care must be taken not to reverse the right side and the left side views Edge views of three reference planes are indicated in Fig 5 For example H R P is the edge view of a horizontal reference plane from which all vertical height dimensions of the object may be measured Similarly all depth dimensions are measured from the frontal reference plane F R P the edge view of which is seen in the two side views and in the top view Width dimensions are measured in the top and front views from the profile reference plane P R P

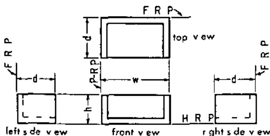


Fig 5 Principal views of reference planes (From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw Hill 1959)

Reference planes Figure 6 shows the front and top views and four auxiliary elevations of a bearing In the front view a horizontal reference plane is taken through the axis of the bearing The same plane is taken in the auxiliary elevations The height dimension or elevation of each point in the front view above or below the reference plane is measured in the front view and then is transferred to each of the auxiliary elevations Because planes are two-dimensional the edge view of a reference plane never should be visualized as a line nor should the end view of a line ever be visualized as a *point*

In practice views are drawn from orientations that show the true sizes of selected lines or surfaces that is views are drawn perpendicular to such lines or surfaces Thus a triangular pyramid might be viewed from four directions each of which shows one face in its true shape

Reading the drawing The object in a drawing is regarded as stationary As the reader looks at first one view and then another he gradually obtains a detailed mental picture of the object The reader imagines that he moves around the object viewing it successively in each direction indicated by the view Each view is regarded as standing out from

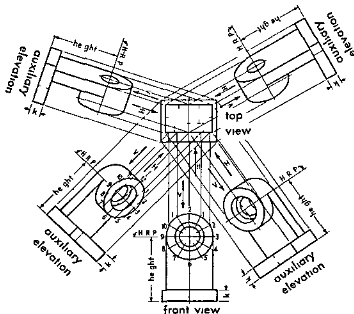


Fig 6 Auxiliary elevations of a bearing (From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw Hill 1959)

the paper, no view is regarded as flat or as a combination of lines on paper, each is a view of the three dimensional object

Geometrical elements of structures. Points, lines and surfaces are the geometrical elements of structures Points are elements of lines, and lines are elements of surfaces Lines are generated by a moving point surfaces are generated by a moving line The law that governs the motion of the moving point or line determines the nature of the generated line or surface In the accompanying chart, various kinds of lines and surfaces are named

Geometrical elements of structures

- Points
- Lines
  - Straight
  - Curved
    - Single curved
      - Circle
      - Ellipse
      - Parabola
      - Hyperbola
      - Trochoid
      - Spiral
      - Involute
      - Cycloid
      - Epicycloid
      - Hypocycloid
      - Sinusoid
    - Double curved
      - General type
      - Helix
- Surfaces
  - Ruled
    - Plane
      - Triangle
      - Quadrilateral
      - Polygon
      - Prism
      - Wedge
      - Pyramid
      - Regular polyhedron
    - Single curved
      - Cylinder
      - Cone
      - Convolute
    - Warped
      - General type
      - Hyperbolic paraboloid
      - Conoid
      - Helicoid
      - Hyperboloid of one sheet
      - Cylindroid
  - Double curved
    - General type
    - Sphere
    - Ellipsoid
    - Paraboloid
    - Hyperboloid of two sheets
    - Torus
    - Surfaces of revolution
    - Serpentine

The first step in preparing a descriptive diagram is to decide what views are necessary to show the geometrical relations and locations of all elements and parts of the object To draw the views, it is necessary to understand the measurements that must be made in each view, the principles involved in showing parallel perpendicular and angular relations between elements, the methods of showing the true lengths of lines and the true shapes of plane surfaces, the intersections of surfaces, and the development of surfaces

True length of line A normal view of a line or plane is taken in a direction perpendicular to the line or plane (Fig 7) The true length of the line AB is seen in the normal view, here taken in a horizontal direction that is perpendicular to the line AB in the top view Dimension h is obtained from the front view The inclination of line AB to the horizontal also is indicated

Parallel lines Parallel lines appear parallel in every view as is illustrated by a parallelogram (Fig 8)

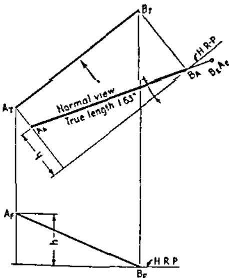


Fig 7 True length of a line (*From G J Hood and A S Palmerlee, Geometry of Engineering Drawing 4th ed McGraw Hill 1959*)

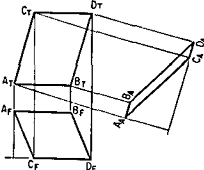


Fig 8 Parallel lines (*From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw Hill, 1959*)

Perpendicular lines Perpendicular lines appear perpendicular only in a view that is normal to one or both of the lines (Fig 9) In Fig 9 RC is a given line, but only the top view of the line RB is given The front view of RB is to be found The numerals indicate the consecutive order in which the lines are drawn A normal auxiliary elevation is

drawn perpendicular to the line RC in the top view In this normal view line RC is drawn and RB is drawn perpendicular to RC This view now shows a normal view of RC but not of RB The front view of RB may now be drawn

Angle between lines The angle between lines AB and AC is to be found (Fig 10) An auxiliary elevation taken in the direction of horizontal line BE is drawn This is an edge view of the plane containing given lines AB and AC Next a view taken in a direction perpendicular to this plane shows a normal view of the two given lines and the true size of the angle between them

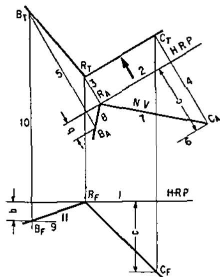


Fig 9 Perpendicular oblique lines (From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw Hill 1959)

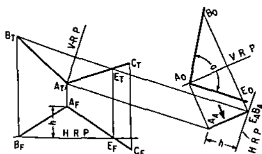


Fig 10 Angle between lines (From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw Hill 1959)

Intersection and development of surfaces A structure is limited by the surfaces that bound it Each surface of the structure is in turn limited by the straight or curved lines in which it intersects adjacent surfaces These lines of intersection determine the edges and joints of the structure and are shown in drawings to describe the structure In addition to the lines of intersection the outlines or contours of curved surfaces also are shown in drawings All intersections must be accurately located to make accurate developments

For example sheet metal bends or folds along a straight line hence only plane-faced and single curved surfaces may be developed The single curved surfaces are cylinders cones and convolutes Warped surfaces and double curved surfaces cannot be developed but must be formed by stretching the metal Each type is useful in engineering designs

A four piece elbow is constructed from parts of right circular cylinders of uniform cross section (Fig 11) At the right the development of each section is outlined by the solid lines Allowance for seams is indicated by dashed lines No material is wasted

A turbine casing is an elbow of many parts and of gradually increasing cross section (Fig 12) Developments of such structures must be drawn accurately for parts to fit closely when assembled

The warped surfaces such as the hyperbolic paraboloid the conoid the helicoid hyperboloid of revolution and others have definite uses and are relatively easy to construct if the geometry of the surfaces is understood

Topographical problems Engineers encounter topographical problems dealing with land contours underground and surface workings and surveys highways bridge piers dams foundations retaining walls excavations underground and surface water supplies drainage irrigation geological formations rock strata oil wells mines tunnels

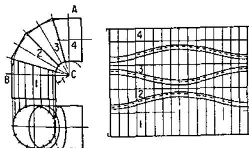


Fig 11 A four piece elbow (From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw-Hill 1959)

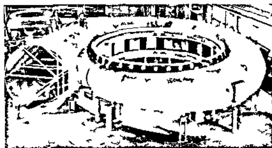


Fig 12 Turbine casing is a development (From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw Hill 1959)

veins of ore dumps layout of grounds landscaping and structures built into the ground and on the surface of the ground (Fig 13)

Fig 13 Veins and outcrops (*From G J Hood and A S Palmerlee Geometry of Engineering Drawing 4th ed McGraw Hill 1959*)

The upper half of Fig 13 illustrates the map of a section of mountainous country The section has been prospected and outcrops or traces of veins of ore have been located From the data given items such as the strike dip thickness line of outcrop and overburden of the veins can be determined and the location and length of tunnels and shafts established [G J H A S P]

Bibliography G J Hood and A S Palmerlee *Geometry of Engineering Drawing* 4th ed 1958 E C Pare R O Loving and I L Hill *Descriptive Geometry* 2d ed 1959 B L Wellman *Technical Descriptive Geometry* 2d ed 1957

Desert

An area characterized by extreme aridity and scanty growth of xerophytic or drought resistant vegetation There are two kinds of desert structure one composed of mountains and hills separated by dry valley flats the other of rocky windswept plateaus and wide but shallow sand filled basins Desert plant life is collectively called desert shrub It consists of grasses shrubs and other plants adapted to aridity in one of three ways quick growing annuals which go through their cycles on the moisture of a single rain or rainy season and exist through drought as seeds perennials such as grass and shrubs that stay alive as roots even though the ground becomes completely dry and armored succulent plants such as cacti which store water in their tissues and protect it by thick skins thorns and other structures In deserts both human and animal life is tied to oases where water supplies are available *See* OASIS VEGETATION ZONES (WORLD)

Desert vegetation and land-surface character of basin floor and bordering eroded mountains near Mammoth Station Imperial County California Although conspicuous dunes are not predominant in most desert lands Only high bordering mountains induce enough precipitation to become heavily forested (*W C Mendenhall USGS*)

Deserts are climatically defined by various formulae which consider the amount and distribution of rainfall in relation to temperature By any formula desert rainfall is much less than potential evaporation and the soil is dry during much of the year [C M D]

Desert erosion features

A distinctive topography carved by erosion in regions of low rainfall and high evaporation where vegetation is scanty or absent Although rainfall is low it is the most important climatic factor in the formation of desert erosion features Desert rains commonly occur as torrential downpours of short duration with a consequent high percentage of runoff As a result of the dryness wind and mechanical weathering also play an important part in desert erosion *See* WEATHERING PROCESSES *see also* SEDIMENTATION (GEOLOGY)

Erosional agents in deserts The principal agents of erosion in deserts are the atmosphere running water and wind

Weathering involves both mechanical and chemical processes Since chemical processes require moisture mechanical weathering predominates in the desert although chemical action is not altogether lacking Rocks are broken by unequal expansion and contraction of constituent minerals and by unequal heating and cooling of outer and inner layers These processes are aided during the rare periods when there is available moisture by the swelling of some minerals as they become hydrated or oxidized and in some localities by crystallization of wind blown salts in cracks Frost wedging may also take place in the winter months when there is a combination of rare rain with freezing temperatures

When storms of the so called cloudburst type occur in the desert sudden rushes of water or flash floods sweep down the normally dry washes or the narrow canyons in the mountains bordering the

basins The comparatively large volume of water combined with a high velocity due to the steepness of the slopes give the short lived streams power to carry large amounts of fine and coarse rock fragments As a result the streams have great erosive power

Where intermittent streams leave the canyons and spread out at the foot of a desert mountain they lose velocity and quickly drop the coarsest of the transported material to build an alluvial fan Some of the water sinks into the fan and some evaporates but whatever remains may follow one of the channels on the fan or spread out in the form of a sheetflood in either case carrying coarse sand silt and clay and perhaps rolling some larger rock fragments along

When the water reaches the toe of the fan it spreads still more dropping all but the finest silt and clay Any excess water follows shallow washes to the lowest part of the basin where it may form a playa lake This evaporates in a few hours or a few days depositing the silt and clay mixed perhaps with soluble salts The flat surfaced area resulting from this deposition is a playa (*see* PLAYA)

A variation of the action of running water occurs if a large accumulation of completely disintegrated material becomes thoroughly water soaked by a sudden hard rain and moves down a canyon and out on the fan as a mudflow Because of the high viscosity and density of the mass of mud and water large boulders may be carried or rolled considerable distances

The lack of moisture during most of the year and the scanty vegetation make the wind a more potent agent of erosion in deserts than in humid lands The finest material is blown high in the air and may be carried entirely out of the area a process known as deflation The larger sand grains are rolled along the surface bouncing into the air when they strike an obstacle knocking more grains into the air as they hit the ground again until eventually a sheet of sand is moving along in the 3 or 4 ft above the surface This moving sand abrades rocks and other objects with which it comes in contact at the same time the grains themselves become rounded and frosted If movement is impeded by vegetation or other obstacles sand accumulates to form dunes *See* DUNE

Erosion cycle in deserts Some knowledge of the erosion cycle in arid regions seems necessary to an understanding of the formation of the distinctive erosional features of deserts as well as the relationships between them

Youthful stage During initial development the bold mountain ranges in or bordering desert areas become gashed by steep canyons and shed waste into the adjoining basins or lowlands as erosion is accelerated during the infrequent but violent rain storms Alluvial fans are built washes develop playas form and the basins slowly fill with detritus (Fig 1) As this stage progresses some alluvial fans coalesce to form bajadas or piedmont alluvial plains along the mountain fronts and individual basins may become deeply filled with waste to form bolsons Desert flats develop between alluvial fans (or bajadas) and playas and isolated dunes accumulate on the lee sides of the latter If the original highlands are flat topped rather than tilted mountain blocks mesas develop (Figs 2 and 3)

As the mountain fronts slowly retreat under the attack of the atmosphere and running water small bare rock surfaces or pediments form at the canyon mouths the result of lateral cutting by the intermittent streams The pediments increase in size in the late part of the stage The general tendency during youth is for relief to decrease

Mature stage The middle stage is initiated by the development of exterior drainage or the capture of higher basins by lower ones as drainage channels erode headward through low divides (Fig 1c and *d*) The fill deposited during youth undergoes erosion and pediments become more widely developed cut not only on the bedrock of the original mountain blocks but also upon the deposits of the captured basin

The mountains are worn still lower and more and more channels extend completely through them cut by the streams engaged in draining and

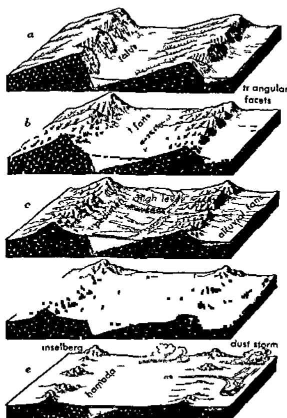


Fig 1 Series of block diagrams (a through e) illustrating a sequence of landforms in an arid climate (Drawn by E Raisz for P E James An Outline of Geography Ginn 1935)



Fig 2 Representative oblique view of mountain and bolson desert in Death Valley California Note at left the beginnings of pediments being developed in the widening mouths of the mountain valleys (California Division of Mines)



Fig 3 Present desert floor in foreground with remnants of higher structures in background From left small mesa pinnacle remnant butte (W T Lee USGS)

dissecting the higher basins Playa deposits or other easily eroded sediments are cut into bad lands before being entirely removed and mesas are reduced to buttes Undissected remnants of older deposits become covered with desert pavement Where winds are turbulent and large supplies of sand are available complex dune areas or even great ergs develop Relief shows some net increase during maturity

Stage of old age The original mountains are so reduced in elevation that the winds sweep over them with little or no condensation of moisture and rains become still more infrequent Great expanses of wind scoured bare rock or hamada are exposed with here and there a more resistant remnant standing above the general level as an *inselberg* (Fig 1e) Buttes are reduced to bornhardts and finally disappear

Those parts of the flat surface floored by earlier deposits are covered and protected by extensive areas of desert pavement The rock fragments may be colored brown to black by desert varnish a coating of manganese and iron oxides

Sand blown from the bare rock surfaces and from the sediments may form large dune areas If there are no obstacles to obstruct movement or cause wind turbulence the sand may move as a sheet forming large expanses of flat or gently undulating sand surfaces Relief slowly decreases during old age

The final result of desert erosion as of erosion under humid conditions, is the peneplain While such a surface is theoretically possible it is doubtful that one has been attained anywhere during recent geologic time See FLUVIAL EROSION CYCLE.

Desert physiographic features There are a number of physiographic features characteristic of desert erosion and deposition

Alluvial fan Where intermittent streams flow down steep canyons in mountains bordering desert areas alluvial fans are formed As the streams suddenly lose velocity on emerging from the canyons at the mountain front they drop most of their load building a fan shaped deposit Such fans consist of a rudely cross bedded mixture of coarse and fine rock fragments largely subangular

Badlands The intricate dissection of relatively fine grained more or less horizontally bedded poorly consolidated sediments results in badlands which are characterized by sharp-edged sinuous ridges separated by steep-sided narrow winding gullies

Bajada A bajada is formed as the result of lateral growth of adjacent alluvial fans until they finally coalesce to form a continuous deposit along a mountain front

Bolson A desert basin of interior drainage which is almost filled by waste from the surrounding mountains is called a bolson

Bornhardt The last remnant of a once-elevated area a bornhardt is reduced to small dimensions by almost equal backweathering on all sides

Butte The erosion under arid conditions of a flat topped surface of soft sediments protected by a resistant cap forms a relatively small remnant of a few acres called a butte Its sides are steep approaching the vertical and may be some hundreds of feet high

Deflation Fine material is blown completely out of a desert region by wind an erosive process called deflation which results in lowering of the surface

Desert flat A large part of a desert area may consist of desert flats which are essentially level surfaces extending from the edge of a playa to the alluvial fans or bajadas bordering a basin

Desert pavement When the finer particles have been removed by deflation, the coarser materials form a desert pavement This mosaic of flat lying interlocking angular to subrounded stones is left as a protective covering over the remainder of the fine material on the desert floor

Dry wash The bed of an intermittent stream in arid or semiarid regions a dry wash (wadi) is generally flat bottomed, its sides are usually vertical ranging from a few feet in height to over 100 ft Arroyo is the term used for a deeper wash

Hammada Ordinarily a hammada is a bare rock surface composed of relatively flat lying consolidated sedimentary rocks from which overlying softer sediments have been stripped principally

by wind erosion Hammadas may also be extensive surfaces cut on bedrock by protracted desert erosion

Inselberg A resistant remnant of a former mountain mass rising above the general level of an almost flat bare rock surface an inselberg is formed by stream planation An outlying peak almost buried by alluvial deposits is sometimes called an inselberg

Mesa A large flat topped surface with an area of a few to many square miles and bounded by steep to nearly vertical sides is called a mesa

Pediment A pediment is a piedmont slope much like a bajada but formed from a combination of processes which are mainly erosional The surface is chiefly bare rock but may have a covering veneer of alluvium or gravel A pediment may be formed at the mouth of a canyon by lateral cutting of an intermittent stream as it swings back and forth seeking new channels at the head of an alluvial fan the cutting may be aided by sheet wash A pediment also may be formed when gradients have been reduced by the filling of a basin or when a higher basin is captured either by a lower basin or by exterior drainage Deposits of the captured basin are beveled by headward erosion of the capturing drainage One type of pediment may grade into the other

[T C]

Bibliography T Clements et al *A Study of Desert Surface Conditions* U S Army Tech Rept EP 53 1957

Desiccant

A substance used to withdraw moisture from other materials Although the removal of large quantities of water is done by evaporation aided by moving air currents and by elevated temperature the last traces of moisture are often held very tightly and do not evaporate readily Furthermore evaporation ceases when the moisture content of the material is reduced to that of the drying air current For final drying one uses as a desiccant a substance with high affinity for water It may react with water chemically or retain water through capillarity or adsorption The drying agent is placed directly into the gas or liquid to be dried solid materials are placed in a desiccator a closed vessel in which moisture diffuses to the desiccant through the dry desiccator atmosphere A desiccant loses potency as it takes on water often it can be renewed by heating Desiccants which form hydrates can be selected to maintain certain levels of low humidity in a closed vessel [A L H]

Important desiccants Silica gel possesses a high adsorptive power because of its extreme capillarity the capillary pores occupying approximately 50% of its specific volume The capillaries are probably spine shaped and the average pore diameter has been estimated to be 4×10^{-7} cm which is only about 10 times the diameter of one molecule of adsorbate The drying efficiency of silica gel depends upon the concentration of water in the gas mixture the temperature of the gel and gas the properties of the condensed liquid its wettability and the state of the gel itself Silica gel adsorbs water vapor preferentially in the presence of other vapors It is readily capable of drying air to a dew point below −94°F

Activated alumina is prepared from aluminum trihydrate It is a granular porous adsorbent with properties similar to those of silica gel Alumina gel has many applications in addition to gas drying It is used to adsorb gases and vapors from gaseous mixtures and to dry liquids

Anhydrous calcium sulfate or Drierite is prepared from a high grade of gypsum $CaSO_4 \cdot 2H_2O$ which is dried crushed sized and heated to 450–500°F for 2 hours This leaves a granular porous form of anhydrous calcium sulfate with sufficient mechanical strength to support its own weight

Magnesium perchlorate Anhydrone is the equal of any desiccant from the standpoint of drying efficiency The adsorption rate is rapid and the first hydrate does not lose water until 275°F is reached thereby permitting its use for drying gases at higher temperatures than most commercial desiccants It passes through three hydrate stages namely di tri and hexahydrate The latter $Mg(ClO_4)_2 \cdot 6H_2O$ represents saturation after adsorption of 48.6% of the dry weight of water a high capacity Although drying efficiency tends to decrease after each hydrate formation even the trihydrate is superior to solid NaOH and $CaCl_2$

Other solid desiccants that have been used or studied for gas drying are oxides such as barium and calcium oxide and activated carbon Barium oxide maintains a high drying activity up to 1000°F Barium oxide also appears to have marked possibilities for the drying of gases at high temperatures

Calcium oxide has long been used as a desiccant because of its low cost Although it has high drying efficiency its capacity is low because of the formation of carbonates on its surface from the carbon dioxide of the air

Activated carbon is old historically and probably has been as intensively studied as any single adsorbent Although capable of adsorbing large amounts of water vapor activated carbon finds its major use in solvent recovery in odor and taste removal and as a catalyst and catalyst carrier Instead of the preferential adsorption of water vapor as in the case of alumina and silica gel organic vapors tend to displace any water present on the carbon *See* ADSORPTION DELIQUESCENCE DRYING

[W R M]

Design standards

Generally accepted uniform procedures dimensions materials or parts that directly affect the design of a product or facility A design standard may be a widely held convention such as the specification of linear dimensions in feet and inches or a particular manufacturer s practice such as the consistent use of a particular size screw to fasten

covers The value of a standard lies in the economies that it produces

Many design standards arise in situations where engineering considerations are insufficient to establish one best design Thus the required strength for a bolted joint could be achieved by many combinations of size and number of bolts Another consideration that leads to design standards is the effect of slight changes in requirements A second bolted joint may be required to carry a slightly greater load than the first The second joint could contain the same number of bolts as the first but use bolts of a slightly larger size or it could contain one more bolt of the previous size Because experience shows that manufacturing economies are greater from procuring bolts of the same size than from fabricating joints with the same number of bolts a limited number of bolt sizes with finite gradations are made standard for design purposes

Benefits of standardization The adoption of standard methods and materials assures that new designs benefit from past experience This aspect of engineering gives rise to governmental regulations and commercial practices covering allowable stresses safety features methods of design symbols to represent materials shapes and finishes and procedures for the manufacture rating inspection and maintenance of products

Standardization enables the designer to communicate a detailed specification in relatively compact form it simplifies procurement and production by the elimination of unnecessary variations and sizes For economy one should specify as few different

signer calculates the size of part theoretically required for a product adds a safety factor established by experience which is itself a form of design standard and selects the next larger standard size part The great variety of gears based on relatively few types of gear teeth pressure angles pitches and the corresponding cutters illustrate the simplification that is achieved by such standardization

In another sense standardization enables an engineer to anticipate the chain of interacting factors that affects a design standardization stabilizes the design environment so that the engineer can move with assurance Standards for grades of materials assure him of uniform properties upon which to base his calculations Standards for manufacturing processes assure him that the product he specifies can be produced Standards for acceptance tests assure him that experience accumulated on previous designs is applicable to the present one

Standardization includes elementary parts surfaces materials processes tools machines methods of test and even the form in which specifications are presented Standardization *reduces to* routine as many phases of engineering as possible thus freeing the engineer to devote himself to the unique features of each project Normal and recurrent situations are the proper domain of standardization exceptional cases where one can show objectively that a departure from standards will improve a product commensurate with the special effort are beyond the realm of standardization

However much engineering can be avoided by adhering to applicable standards The success of a new product may depend on its conformance to industry or customer standards A major engineering function is to seek out and call attention to existing applicable standards Standardization produces real and far reaching economies

Situations suitable for standardization Ball bearings because of their widespread use are the subject of international standards Screw threads are the subject of nation wide standards and through the International Organization for Standardization are achieving international status Hoisting rope is standardized within one segment of industry A company establishes internal standards for drafting room practice fabricated parts stock parts and materials using industry wide standards such as the American Drafting Standards Manual where possible

The more repetitive a manufacturing process is the greater the opportunity for standardization The more a line of products can be assembled from similar elements as in the assembly of electronic circuits from resistors inductors and capacitors the more the elements can be standardized In this respect standardization assures equipment manufacturers of a ready source of parts by enabling vendors to anticipate requirements and assists parts vendors in obtaining a stable market by encouraging equipment manufacturers to purchase like parts Advantages of standardization in reducing cost simplifying replacement and decreasing the quantities of materials carried in stock become more apparent as manufacturing volume increases

To the creative engineer unacquainted with tooling cost inventory management marketing field service and maintenance the limited variations imposed by standardization may seem a hindrance However standardization on particular shapes of structural parts for example enables the parts to be produced by extrusion or rolling instead of by more costly machining operations Products can be designed so that some parts are interchangeable between models The designer looks for opportunities to use parts for which patterns or dies are already available instead of designing a new part that serves the purpose no *better*

Introducing new standards Most engineering societies and trade associations develop standards and bring them up to date as a major part of their activity Branches of the government especially those engaged in procurement have standardization groups In the United States some of their efforts are correlated through such agencies as the Federal Specifications Board Large companies and especially manufacturers that operate several widely separated plants provide a separate stand

ards group to seek out subjects suitable for standardization

General engineering standards such as those promulgated by the American Standards Association are formulated over the years by committees and by circulation to a representative segment of the interested industry As a consequence an adopted standard embodies far more critical study than an individual designer could justify for a single design It is not likely that one man s modification will represent a significant improvement over such a thoroughly analyzed solution For example standards for cutting tools are the result of much experience and controlled tests Profitable operation of high speed automatic machine tools is possible in great measure from this accumulated experience

Industry sponsored organizations such as the Underwriters Laboratory consider product designs from multiple points of view The design recommendations and requirements established by such cooperative organizations reconcile the state of an art with user safety

Standardization in a rapidly developing industry is both difficult and dangerous but adequately basic standards withstand the test of time Materials handling is an especially attractive area for

mizes storage costs material in process inventory damage to goods and handling costs

A standards engineer or group collects information performs tests and develops standards as a service to all operating departments in a company However final authority for issuance of standards should reside with departments that will use and be responsible for compliance with the standards Standard costs are usually derived in the accounting office after materials processes and procedures are themselves standardized Such standardization requires in turn standardized equipment which assumes a stable activity that will continue long enough to enable the investment to be recovered

Design standards originate in the course of an engineer s daily work Possibly part of an existing product is applicable to a new product but could be improved Before redesigning it for the new product the engineer reviews its use in the existing product so that the new version can be used in both products Or in setting up an additional assembly line a methods engineer introduces an improved procedure The procedure is also standardized on existing assembly lines and workers are trained in its advantages and use Thus improvements are carried back to existing products and processes as well as carried forward to new situations as standard designs

Scope and applicability The more diversified the group that cooperates to promulgate a standard the fewer the items that can be included The American Standards Association in conjunction with industrial and professional societies has brought into being a wide variety of standards touching on almost every phase of technology All standards are subject to review and reaffirmation or revision in the light of advances at least once in 5 years

One of the most useful and important standards is that on metal fits yet its use is spreading only slowly The standard on tolerances allowances and gages for metal fits is basic to all manufacturers of metal parts This standard establishes the basic hole specifications That is the nominal diameter of a hole is taken as the base from which other dimensions are derived From this minimum diameter an allowance is subtracted for running fits to give the maximum shaft diameter The allowance is therefore the minimum diametrical clearance provided to enable the shaft to function properly in the hole The hole is given a plus tolerance so that it may be slightly larger than nominal and the shaft is given a minus tolerance so that it may be slightly smaller than nominal In this way any shaft of a nominal size will always pass through any hole of the same nominal size

Fundamentally standardization implies approaching a problem at so basic a level and studying it from so many aspects that the final solution is permanent Scientific principles inhere in the nature of things Standards are based on these principles but add to them elements of experience preferences practices and mutual agreement among cooperating persons [F H R]

Desmodoroidea

A superfamily of marine nematodes with a ringed but smooth cuticle reflexed ovaries and amphids of many shapes Species of one family have hollow bristles containing adhesive glands in front of the anus and utilize these to move about in a leechlike manner These species appear to have a head but this results from the contrast between a swollen anterior and a slender neck region Another family uses the same method of locomotion but has solid bristles Some authorities give this group the status of an order *See* NEMATODA [H F W]

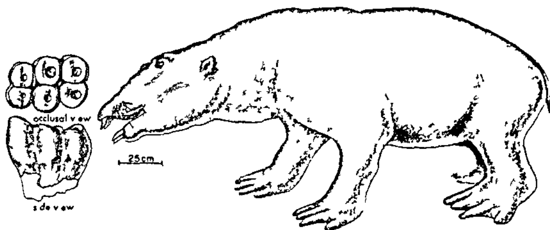
Desmoscolecoidea

A small group of free living nematodes usually considered to be a superfamily characterized by a ringed body an armored head set off from the body and hemispherical amphids The species are *small and plump and with the exception of a species* found in caves in Yugoslavia are marine Two families are recognized Desmoscolecidae and Greefiellidae Members of the Desmoscolecidae have coarse annulation and resemble annelids or small insect larvae those of the Greefiellidae which contains one genus lack annules but have more bristles *See* NEMATODA [H F W]

Desmostylia

An extinct order of large hippopotamuslike amphibious gravigrade shellfish eating

ly 12 ft long with a northern trans Pac



A restorat on and cheek tooth of *Desmostylus* from the M ocene of Cal forn a (*Redrawn from R A St rton 1959*)

tribution They frequented shallow bays and coastal marshes during Oligocene and Miocene time

There are two families the Desmostylidae (*Desmostylus Cornwallius* and *Vanderhoofia*) and the Paleoparadoxidae (*Paleoparadoxia*) The many cusped molars are composed of a cluster of heavy enameled cylindrical columns resembling the pavement teeth in drum fish The cheek teeth are anterovertically replaced as in the Proboscidea and Sirenia There are one to four pairs of procumbent tusks These animals probably descended from an ancestral stock that also gave rise to the Proboscidea and Sirenia *See* PROBOSCIDEA FOSSILS SIRENIA FOSSILS [G T J]

Destructive distillation

The primary chemical processing of materials such as wood coal oil shale and some residual oils from refining of petroleum It consists in heating material in an inert atmosphere at a temperature high enough for chemical decomposition The principal products are (1) gases containing carbon monoxide hydrogen hydrogen sulfide and ammonia (2) oils and (3) water solutions of organic acids alcohols and ammonium salts For a discussion of the products of the destructive distillation or coking of coal *see* COAL CHEMICALS

Crude shale-oil obtained by destructive distillation of carboniferous shales is being produced on a commercial scale in Scotland Latvia and Sweden and on a pilot plant scale in the United States Crude shale-oil may be subjected to a destructive or coking distillation to reduce its viscosity and increase ... quent cat ... alumina ... sure) lo ... that the oil can then be refined by normal petroleum refinery operations Residual oils from petroleum refinery operations are subjected to coking distillation so as to reduce the carbon content The coke is used for the manufacture of electrode carbon and the oil is returned to the feed for normal petroleum refining

Prior to about 1920 destructive distillation of wood was an important source of methanol, acetic acid and acetone Currently these chemicals are produced at a lower cost by other methods for example methanol is produced by hydrogenation of carbon monoxide acetaldehyde and acetic acid from acetylene and acetone by fermentation processes and by decomposition of cumene hydroperoxide The main product of the destructive distillation of wood is 40–45% charcoal used in metallurgical processes in which the low content of ash sulfur and phosphorus is important Special chars made by the destructive distillation of the shells of nuts have very large surface areas and are used as adsorbents in gas masks and in some chemical processes Char from the destructive distillation of bones is used for decolorizing solutions of raw sugar and the oil obtained during the decomposition of bones contains recoverable amounts of pyridine pyrrole quinolines and other nitrogen compounds *See* CHARCOAL COKE COKING (PETROLEUM REFINING) OIL SHALE PYROLYSIS WOOD CHEMICALS [H H ST]

Detector

A nonlinear device employed to recover the desired signal from the modulated wave also called a demodulator In radio communications the signal to be transmitted must first be impressed or modulated upon a periodic wave called the carrier (*see* MODULATION MODULATOR) Upon reception of the modulated signal it is necessary to remove the original signal from the modulated wave by a process known as demodulation or detection Because transmission can be accomplished by employing amplitude frequency or phase modulation the process of detection and the practical circuits for accomplishing it will differ in each case Detection of amplitude modulated periodic waves

is carried out by the process known as rectification which results in pulsating currents whose envelope corresponds to the desired signal Detection of frequency and phase-modulated signals is usually obtained by first passing the modulated wave through a circuit which causes the wave to be converted to an amplitude-modulated wave from which the desired signal is derived by an amplitude modulation detector

Most frequent use of detectors occurs in circuits employed to receive various classes of radio signals such as radio and television Detectors are also used in various classes of measuring systems where it is often desirable to produce an indication of the presence of signal or its relative magnitude as some quantity of interest is varied The indicating devices employed in impedance bridges or standing wave detectors are examples of such measurement applications In some cases, such as an electronic radio-frequency voltmeter the carrier may not be modulated and one may wish to provide a detector to indicate or measure the presence of the carrier

Types of detector All detectors require the use of nonlinear devices for removing the signal from the modulated wave The simplest of these are the thermionic diodes crystal rectifiers and copper and selenium oxide rectifiers Electronic diodes have relatively low internal capacities and are suitable for use at all frequencies up to the microwave region Crystal rectifiers because of their low internal capacities and small size may be used at all common radio and microwave frequencies They have the further advantages of not requiring cathode heating power and being smaller in size and lower in cost For these reasons they are ideal in many applications Copper and other oxide rectifiers are useful at the lower radio frequencies, and have the advantage of a high power handling capacity and the simplicity of a crystal rectifier Nonlinear characteristics of triodes tetrodes and more complicated electronic devices are also useful for special applications Triodes and other multielement vacuum tubes can be used simultaneously as detectors and amplifiers thus permitting considerable circuit simplification due to the high degree of amplification which is possible in regenerative and superregenerative detectors

Detection fidelity In most applications it is important to consider the fidelity of the detection process An ideal linear detector is one which faithfully reproduces the modulation existing upon the carrier and produces an output signal whose magnitude is proportional to the magnitude of the modulation envelope A detector may fail to accomplish this in one or more of the following ways (1) amplitude distortion is said to result if the output of the detector contains frequencies which are not present in the modulation envelope (2) frequency distortion results if the various frequency components of the modulation envelope are reproduced with different amplitudes and (3) phase distortion is said to result if the frequency components of the modulation envelope are reproduced with altered phase relationships

An ideal square detector is one in which the amplitude of the output signal is proportional to the square of the effective value of the carrier amplitude Any nonlinear device becomes a square law detector when the applied signals are sufficiently small By restricting the amplitude of the incoming signal the response law becomes precisely known for this reason square law detectors are useful in measurement applications

See AMPLITUDE MODULATION DETECTOR FREQUENCY MODULATION DETECTOR PHASE MODULATION DETECTOR [E L C]

Bibliography J F Reintjes and G T Coate *Principles of Radar* 3d ed 1952 F E Terman *Electronic and Radio Engineering* 4th ed 1955

Detergent

A substance used to enhance the cleansing action of water Because oily films are primarily responsible for the attachment of dirt particles the role of the detergent is that of wetting agent and emulsifier It reduces the interfacial tension between water and oil this allows the particle to become wet and to float off Emulsifying agents are molecules having an oil like nonpolar portion combined with a polar group the former is readily drawn into oil and the latter into water thus forming a bridge between the two As a result the oil is broken into tiny droplets that is it forms an emulsion

Soap the sodium salt of long chain acids was the principal detergent until superseded in production in 1954 by synthetic detergents These synthetic detergents are of three types anionic for example the sodium salts of medium chain length (7 18 carbons) alkyl sulfates or sulfonates cationic for example the tetraalkyl ammonium halides which are costly but effective metal cleaners and nonionic for example products made from tall oil a paper industry by product by reaction with ethylene oxide to form low foaming esters that are valued for use in automatic washers Commercial detergents often contain tetrasodium pyrophosphate or other builders which enhance cleansing *See* SOAP AND DETERGENT SURFACE ACTIVE AGENT [A L H]

Determinant

The concept of a determinant can best be explained with reference to a matrix A

$$\text{Let} \qquad A = \begin{pmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{pmatrix}$$

be an n by n square array of numbers Such an array is called an n by n square matrix and the numbers a_{ij} in the array are called elements of the matrix Determinants are valuable in the solution of sets of linear equations The determinant of the matrix A denoted by

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \qquad (1)$$

is a number which is the value of a certain function of the elements of A. Before defining the determinant $|A|$, it may be noted that the theory of determinants had its origin in the solution of linear systems of equations and that determinants occur in virtually all branches of mathematics and related fields (see LINEAR SYSTEMS OF EQUATIONS; MATRIX THEORY; POLYNOMIAL SYSTEMS OF EQUATIONS).

In the linear system of equations

$$a_{11}x + a_{12}y = b_1 \qquad a_{21}x + a_{22}y = b_2 \qquad (2)$$

successive elimination of the unknowns x and y yields a simpler equivalent system

$$\begin{aligned} (a_{11}a_{22} - a_{12}a_{21})x &= b_1a_{22} - a_{12}b_2 \\ (a_{11}a_{22} - a_{12}a_{21})y &= a_{11}b_2 - b_1a_{21} \end{aligned} \qquad (3)$$

The coefficients of equations (3) are determinants

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$\begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix} = b_1a_{22} - a_{12}b_2$$

$$\begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix} = a_{11}b_2 - b_1a_{21}$$

of 2 by 2 matrices with elements which are the coefficients of Eqs. (2).

If $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$ then the solution of (2) is readily given by ratios of the above determinants. Similarly in the system

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= b_1 \\ a_{21}x + a_{22}y + a_{23}z &= b_2 \\ a_{31}x + a_{32}y + a_{33}z &= b_3 \end{aligned}$$

the determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$
$$- a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} \qquad (4)$$

enters into the solution of the system.

The determinant (1) is called a determinant of order n, and it can be defined by induction on n. For explanatory purposes, determinants of orders $n = 1, 2$, and 3 are defined as follows. For $n = 1$, $|a_{11}| = a_{11}$. For $n = 2$,

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}|a_{22}| - a_{12}|a_{21}| = a_{11}a_{22} - a_{12}a_{21}$$

For $n = 3$,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$
$$= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$
$$= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31})$$
$$+ a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

The determinants of order 2 and 3 are defined in terms of determinants of order 1 and 2, respectively. Suppose now that a determinant of order $n - 1$ has been defined. In (1), denote by A_{1j}, $j = 1, 2, \ldots, n$, the determinant of order $n - 1$ formed by deleting from (1) the first row and the jth column. Then by definition

$$|A| = a_{11}A_{11} - a_{12}A_{12} + a_{13}A_{13} + \cdots$$
$$+ (-1)^{n-2}a_{1,n-1}A_{1,n-1} + (-1)^{n-1}a_{1n}A_{1n} \qquad (5)$$

This definition is seen to agree with the definitions for determinants of order 2 and 3, and by the principle of mathematical induction, defines a determinant of order n for every n. There are many other definitions of a determinant of order n which can be proved to be equivalent to the definition given here.

As an example of the definition (5),

$$|A| = \begin{vmatrix} -2 & 0 & 1 & 3 \\ 5 & 2 & -6 & 1 \\ 0 & 4 & -1 & 7 \\ 10 & 2 & 3 & 1 \end{vmatrix} = (-2)\begin{vmatrix} 2 & -6 & 1 \\ 4 & -1 & 7 \\ 2 & 3 & 1 \end{vmatrix}$$
$$- (0)\begin{vmatrix} 5 & -6 & 1 \\ 0 & -1 & 7 \\ 10 & 3 & 1 \end{vmatrix} + (1)\begin{vmatrix} 5 & 2 & 1 \\ 0 & 4 & 7 \\ 10 & 2 & 1 \end{vmatrix}$$
$$- (3)\begin{vmatrix} 5 & 2 & -6 \\ 0 & 4 & -1 \\ 10 & 2 & 3 \end{vmatrix} \qquad (6)$$

$$\begin{vmatrix} 2 & -6 & 1 \\ 4 & -1 & 7 \\ 2 & 3 & 1 \end{vmatrix}$$
$$= (2)\begin{vmatrix} -1 & 7 \\ 3 & 1 \end{vmatrix} - (-6)\begin{vmatrix} 4 & 7 \\ 2 & 1 \end{vmatrix} + (1)\begin{vmatrix} 4 & -1 \\ 2 & 3 \end{vmatrix}$$
$$= 2(-1 - 21) + 6(4 - 14) + (12 + 2) = -90$$

$$\begin{vmatrix} 5 & 2 & 1 \\ 0 & 4 & 7 \\ 10 & 2 & 1 \end{vmatrix}$$
$$= (5)\begin{vmatrix} 4 & 7 \\ 2 & 1 \end{vmatrix} - (2)\begin{vmatrix} 0 & 7 \\ 10 & 1 \end{vmatrix} + (1)\begin{vmatrix} 0 & 4 \\ 10 & 2 \end{vmatrix}$$
$$= 5(4 - 14) - 2(0 - 70) + (0 - 40) = 50$$

$$\begin{vmatrix} 5 & 2 & -6 \\ 0 & 4 & -1 \\ 10 & 2 & 3 \end{vmatrix}$$
$$= (5)\begin{vmatrix} 4 & -1 \\ 2 & 3 \end{vmatrix} - (2)\begin{vmatrix} 0 & -1 \\ 10 & 3 \end{vmatrix} + (-6)\begin{vmatrix} 0 & 4 \\ 10 & 2 \end{vmatrix}$$
$$= 5(12 + 2) - 2(0 + 10) - 6(0 - 40) = 290$$

Hence

$$|A| = (-2)(-90) + (1)(50) - (3)(290) = -640$$

It can be proved from the definition (5) that the determinant $|A|$ is the sum of all terms of the form $(-1)^j a_{1j_1}a_{2j_2} \cdots a_{nj_n}$ for all possible orderings $j_1, j_2, \ldots, j_n$ of the second subscripts $1, 2, \ldots, n$, and the number j is the number of interchanges of two digits required to carry the ordering $j_1, j_2, \ldots, j_n$ into the natural ordering $1, 2, \ldots, n$. Thus $|A|$ is the sum of all products of n elements, one from

each row and one from each column with a certain sign affixed to each product Since there are $n! = 1\ 2\ 3 \quad (n-1)\ n$ orderings of the numbers 1 2 , n there are $n!$ terms in the sum

In the definition of a determinant (5) A_{1j}, $j = 1$ 2 n which is a determinant of order $n-1$ is called the complementary minor of the element a_{1j} of the first row In general the complementary minor of any element a_{ij} is the determinant of order $n-1$, A_{ij}, which is formed from A by deleting the ith row and jth column The cofactor of the element a_{ij} is $(-1)^{i+j}A_{ij}$ Thus in definition (5) the value of $|A|$ is given as the sum of the products of each element of the first row and its cofactor It can be shown that for any row i $|A| = a_{i1}(-1)^{i+1}A_{i1} + a_{i2}(-1)^{i+2}A_{i2} + \quad + a_{in}(-1)^{i+n}A_{in}$. This expansion of $|A|$ is called the expansion of $|A|$ according to the elements of the ith row Thus in example (6)

$$|A| = (0)(-1)^{3+1}\begin{vmatrix}0&1&3\\2&-6&1\\2&3&1\end{vmatrix} + (4)(-1)^{3+2}\begin{vmatrix}-2&1&3\\5&-6&1\\10&3&1\end{vmatrix} + (-1)(-1)^{3+3}\begin{vmatrix}-2&0&3\\5&2&1\\10&2&1\end{vmatrix} + (7)(-1)^{3+4}\begin{vmatrix}-2&0&1\\5&2&-6\\10&2&3\end{vmatrix} = -640$$

is the expansion of $|A|$ according to the third row

The transpose of the square matrix A is the matrix denoted by A' which has as its ith row the ith column of A and as its jth column the jth row of A for all i,j It can be proved that $|A'| = |A|$ It follows from this property that any theorem concerning the rows of a determinant implies a corresponding result concerning the columns and vice versa For example a determinant can be evaluated by an expansion according to the jth column by which $|A|$ is the sum of the products of each element of the jth column and its cofactor

The following further properties of determinants follow from the definition (5)

(i) If B is a matrix obtained from A by interchanging two rows (columns) of A then $|B| = -|A|$

(ii) If two rows (columns) of A are identical then $|A| = 0$

(iii) If B is a matrix obtained from A by multiplying every element of one row (column) of A by the number m then $|B| = m|A|$

(iv) If B is a matrix obtained from A by adding to each element of a row (column) of A a constant multiple of the corresponding element of another row (column) of A then $|B| = |A|$

(v) If B is a matrix identical with A except possibly for the ith row b_1 b b_n and if C is a matrix identical with A except that the ith row of C is $a_1 + b_1$ $a_2 + b_2$ $a_n + b_n$ then $|A| + |B| = |C|$ (There is the corresponding property for columns)

The above properties are used to shorten the computation in finding the value of a determinant In example (6)

$$|A| = \begin{vmatrix}-2&0&1&3\\5&2&-6&1\\0&4&-1&7\\10&2&3&1\end{vmatrix}$$

$$= 2\begin{vmatrix}-2&0&1&3\\5&1&-6&1\\0&2&-1&7\\10&1&3&1\end{vmatrix} \text{ by (iii) applied to the second column}$$

$$= 2\begin{vmatrix}-2&0&1&3\\5&1&-6&1\\-10&0&11&5\\10&1&3&1\end{vmatrix} \text{ by (iv) multiplying the second row by } -2 \text{ and adding to the third row}$$

$$= 2\begin{vmatrix}-2&0&1&3\\5&1&-6&1\\-10&0&11&5\\5&0&9&0\end{vmatrix} \text{ by (iv) multiplying the second row by } -1 \text{ and adding to the fourth row}$$

$$= 2\,(1)(-1)^{2+2}\begin{vmatrix}-2&1&3\\-10&11&5\\5&9&0\end{vmatrix} \text{ by the expansion according to the elements of the second column}$$

$$= 2\begin{vmatrix}0&1&0\\12&11&-28\\23&9&-27\end{vmatrix} = 2(1)(-1)^{1+2}\begin{vmatrix}12&-28\\23&-27\end{vmatrix}$$

$$= -8\begin{vmatrix}3&-7\\23&-27\end{vmatrix}$$

$$= -8(-81+161)$$

$$= -640 \text{ by (iv) (iii) and definition (5)}$$

If A and B are square matrices of order n then the product AB is a square matrix of order n and the element in the ith row and jth column for all $i\,j$ is obtained by multiplying the n elements in the ith row of A into the n elements in the jth column of B term by term and adding these products A most useful property of determinants is the fact that the determinant of the product AB is equal to the product of the determinant of A and the determinant of B that is $|AB| = |A|\,|B|$

In the matrix A select any r rows and r columns ($r \leqq n$) The elements common to these rows and columns form an r by r matrix M and the determinant $|M|$ is called an r rowed minor of A The determinant of the $n-r$ by $n-r$ matrix of elements common to the remaining $n-r$ rows and columns of A is called the complementary minor of $|M|$ When $r = 1$ and row i and column j are selected then $|M| = a_{ij}$ and the complementary minor is A_{ij} which was defined earlier The Laplace expansion of $|A|$ gives the value of $|A|$ as a sum with certain signs of all possible minors formed from a fixed set of r rows multiplied by their complementary minors If i_1 i_2 i_r are the fixed rows then the term in the sum which is the minor formed from the columns j_1 j_2 j_r multiplied by its complementary minor, has the sign $(-1)^{I+J}$ where $I = i_1 + i_2 + \quad + i_r$ and

$J = j_1 + j_2 + \quad + j$ In the example (6) choosing the second and third rows as the fixed rows gives

$$
\begin{aligned}
|A| = {} & (-1)^{2+3+1+2}\begin{vmatrix}5 & 2\\0 & 4\end{vmatrix}\begin{vmatrix}1 & 3\\3 & 1\end{vmatrix}\\
& + (-1)^{2+3+1+3}\begin{vmatrix}5 & -6\\0 & -1\end{vmatrix}\begin{vmatrix}0 & 3\\2 & 1\end{vmatrix}\\
& + (-1)^{2+3+1+4}\begin{vmatrix}5 & 1\\0 & 7\end{vmatrix}\begin{vmatrix}0 & 1\\2 & 3\end{vmatrix}\\
& + (-1)^{2+3+2+3}\begin{vmatrix}2 & -6\\4 & -1\end{vmatrix}\begin{vmatrix}-2 & 3\\10 & 1\end{vmatrix}\\
& + (-1)^{2+3+2+4}\begin{vmatrix}2 & 1\\4 & 7\end{vmatrix}\begin{vmatrix}-2 & 1\\10 & 3\end{vmatrix}\\
& + (-1)^{2+3+3+4}\begin{vmatrix}-6 & 1\\-1 & 7\end{vmatrix}\begin{vmatrix}-2 & 0\\10 & 2\end{vmatrix}\\
= {} & (20)(-8) - (-5)(-6) + (35)(-2)\\
& + (22)(-32) - (10)(-16) + (-41)(-4)\\
= {} & -640
\end{aligned}
$$

The Laplace expansion is particularly useful in evaluating the determinants of matrices which have blocks of zeros for example using the first three rows as the fixed rows

$$
\begin{vmatrix}
-5 & 4 & -1 & 0 & 0 & 0\\
2 & 6 & 3 & 0 & 0 & 0\\
0 & 1 & -1 & 0 & 0 & 0\\
0 & 0 & 0 & 4 & -1 & 0\\
0 & 0 & 0 & 2 & 7 & 1\\
0 & 0 & 0 & 0 & 5 & 3
\end{vmatrix}
= \begin{vmatrix}-5 & 4 & -1\\2 & 6 & 3\\0 & 1 & -1\end{vmatrix}\begin{vmatrix}4 & -1 & 0\\2 & 7 & 1\\0 & 5 & 3\end{vmatrix} = (51)(70) = 3570
$$

Every minor formed from the first three rows is zero by property (II) except the one formed from the first three columns See EQUATIONS THEORY OF

[R A B]

Bibliography R R Middlemiss *College Algebra* 1952 H W Turnbull *The Theory of Determinants Matrices and Invariants* 1928

Detonator

A device used to initiate the explosion of a high explosive See EXPLOSION AND EXPLOSIVE FUSE EXPLOSIVE Detonators employ a very sensitive primary explosive which detonates readily when set afire by the primer Primary explosives in common use are mercury fulminate and lead azide In addition the detonator or blasting cap usually contains a small charge of secondary explosive such as PETN which serves to transmit a more powerful shock wave than the primary explosive alone Detonator caps are sold in a series of numbered sizes such as 4 6 and 8 which contain increasing amounts of explosive

An electric detonator is ignited by a fuse wire which serves to touch off the primer See PRIMER (EXPLOSIVE) The explosion of the detonator readily sets off a dynamite charge The detonator is inserted obliquely into a hole punched in the side of the stick with an awl For less sensitive explosives such as cast TNT however a booster charge is needed This auxiliary charge usually made of pressed tetryl is placed between the cap and the main charge

Alfred Nobel invented the detonator and also discovered safe means for using nitroglycerin (dynamite and blasting gelatin) Nobel s two discoveries gave an epoch making impetus to mining

[W E C]

Deuterium

The isotope of the element hydrogen with atomic weight 2 0147 and symbols H^2 or D Considerations of nuclear stability and a discrepancy between the chemical and physical atomic weights of hydrogen led to the prediction of a stable isotope of hydrogen of mass 2 A successful search for this isotope deuterium was made by H C Urey F G Brickwedde and G M Murphy in 1931 The terrestrial natural abundance of deuterium is 1 part in 6700 parts of ordinary hydrogen (protium) which has atomic weight 1 0089 Small variations in natural sources are found as a result of fractionation by geological processes Industrial hydrogen particularly that generated by the electrolysis of water may contain significantly less deuterium

Deuterium is used mainly in the form of heavy water In the uncombined state it finds uses as a research tool Liquid deuterium is used in bubble chambers to study the reactions of elementary particles with the deuterium nucleus the deuteron Deuterons are frequently accelerated in cyclotrons to study their reactions with other nuclei and also to produce radioactive nuclides See NUCLEAR REACTION Deuterium gas is used in the direct synthesis of organic compounds for tracer studies Illustrative of such procedures are exchange reactions with hydrogen containing substances in the presence of hydrogenation catalysts If controlled thermonuclear fusion can be achieved deuterium gas would become an exceedingly important source of power See FUSION NUCLEAR

Deuterium D_2 is a gas at room temperature It is prepared from heavy water (D_2O) either by electrolysis or by reaction of D_2O with metals such as zinc iron calcium and uranium It is also prepared directly by the fractional distillation of liquid hydrogen In this process it is necessary to catalyze the disproportionation (unsymmetrical dissociation and recombination) of HD into H_2 and D [illegible] of [illegible] At [illegible] equilibrium composition is 97 8% *o* deuterium Deuterium molecules obey the Bose Einstein statistics and the ortho species have even rotational quantum numbers whereas the para species have odd rotational quantum numbers The analysis of the ortho para composition of deuterium is most conveniently made by measurement of the thermal conductivity of the gas at 77°K The physical and chemical properties of *o* and *p* deuterium are very similar In most of its physical

and chemical properties deuterium resembles protium In most cases deuterium is slightly less reactive than protium An intercomparison of some of the physical properties of D_2 with those of H_2 is given in the table

Selected values of some physical properties of deuterium

	n H_2	n-D_2	97 8% o-D_2
Triple point	13 96°K	18 72°K	18 69°K
Normal boiling point	20 4°K	23 6°K	23 5°K
Critical temperature	33 24°K	38.35°K	
Critical pressure	12 8 atm	16 4 atm	
Heat of fusion	28 0 cal/mole	47 0 cal/mole	
Heat of vaporization (at normal boiling point)	216 cal/mole	293 cal/mole	
Molar volumes of liquid at 20°K	28 3 ml	23 5 ml	

Deuterium gas is usually slightly contaminated with HD The analysis of the gas for protium is most conveniently carried out by a mass spectrometer Alternative methods of analysis are by thermal conductivity and the rate of effusion of the gas through an orifice The gas can be converted to water and the following properties have been used as a basis for analysis infrared spectrum density index of refraction and nuclear magnetic resonance

The chemical reactivity of deuterium is less than that of hydrogen because of its lower zero-point energy and smaller collision frequency At 1000°K deuterium is 32% less dissociated than protium At room temperature deuterium atoms are electrolyzed out of water in the form of hydrogen gas at one eighth the rate of protium atoms *See* DEUTERON HEAVY WATER HYDROGEN TRITIUM

[J R]

Bibliography A Farkas *Orthohydrogen Parahydrogen and Heavy Hydrogen* 1935 A H Kimball *Bibliography of Research on Heavy Hydrogen Compounds* 1949 H W Woolley R B Scott and F G Brickwedde Compilation of thermal properties of hydrogen in its various isotopic and ortho-para modifications *U S Natl Bur Standards J Research* 41 397–475 1948

Deuteron

The nucleus of the atom of heavy hydrogen H^2 (deuterium) The deuteron d is comprised of a proton and a neutron As the simplest multinucleon nucleus the deuteron has been the subject of extensive study Its binding energy is 2 227 Mev that is this is the amount of energy which must be added to a deuteron for it to dissociate into a proton and a neutron The accurate determination of this dissociation energy provides the means of calculating the mass of the neutron the mass of the deuteron (2 014187 amu) and proton being known from other experiments

The intrinsic angular momenta or spins of the proton and neutron combine to produce a deuteron spin of unity hence the deuteron obeys the type of quantum statistics known as Bose Einstein statistics The deuteron possesses a magnetic moment (0 857407 nuclear magnetons) and an electric quadrupole moment (2.738×10^{-27} cm^2)

Deuterons are much used as projectiles in nuclear bombardment experiments especially to produce $(d\ p)$ $(d\ n)$ and $(d\ \alpha)$ reactions In the first two reactions because of the low binding energy of the deuteron the neutron n or proton p is stripped from it and captured by the target nucleus Meanwhile the other half of the deuteron (that is the proton or neutron) carries away the excess energy The H^1/H^2 abundance ratio in nature is 6700 *See* DEUTERIUM NUCLEAR REACTION

[H E D]

Deuterostomia

That portion of the animal kingdom which includes the phyla Echinodermata Chaetognatha and Chordata Embryonic development is characterized by indeterminate cleavage of the egg The coelom and mesoderm arise as outpocketings of the gut wall and the mouth is a new structure on the end opposite from the blastopore *See* PROTEROSTOMIA

[T L S]

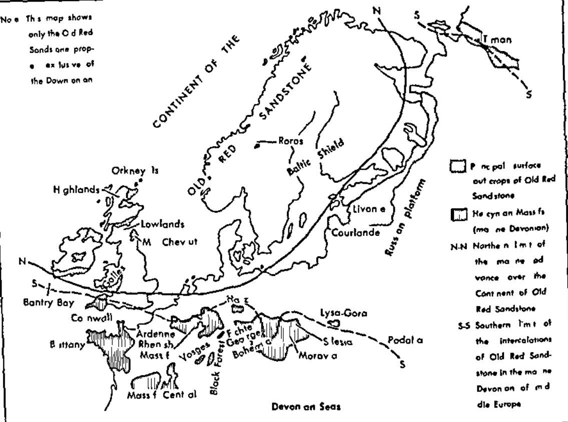
Devonian

The fourth period of the Paleozoic Era The name as originated by R Murchison and A Sedgewick in 1839 was applied to the thick section of marine

PRECAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBONIFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	Mississippian	Pennsylvanian	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

structurally complex fossiliferous rocks in Devonshire southwestern England Much better sections of equivalent age were recognized later in the Rhine Valley and Ardennes The European standard of correlation is based on these sections The excellent section of Devonian rocks in New York State is the standard of reference for North America

The Old Red Sandstone is the continental counterpart of Devonshire marine facies It consists of a thick highly colored succession of alternating conglomerates sandstones and shales of lagoonal and desert origin It is mostly unfossiliferous but remains of fresh water fishes and land plants occur locally These continental deposits are well developed in Cornwall and Wales in Scotland in Scandinavian and Baltic countries in F
in all of northwest Russia to the White



F g 1 D agrammat c map of the cont nent of the Old Red Sandstone and ts border zone (*From M G gnoux translated by G G Woodford Strat graph c Geology Freeman 1955*)

In eastern North America continental facies of the Catskill delta in New York and Pennsylvania are equivalent in time to the uppermost Old Red Sandstone in Great Britain and to the Baltic sandstones of the Russian platform The most complete sequence is in New Brunswick Canada where both lower and upper divisions of continental deposits equivalent to those in Europe are recognized

Comparisons between equivalent continental deposits on both sides of the North Atlantic suggest a continuous land mass connecting Europe and North America (Eria and Laurentia) during the Devonian This probable land bridge has been called the Erian Continent Evidence of the land bridge is found in the similar trend of folded rocks on either side of the Atlantic and in the broken off character of the rocks in each place Similarity of land plants and fresh water animal fossils preserved in Devonian rocks in the two separate regions is indicative of a shallow water connection along an ancient shore line—a connection which permitted easy migration of the then existing forms of plant and animal life The intervening sections of the land mass later subsided beneath the ocean

Sed men f

110 n a shallow bank between Scotland and Greenland may be a remnant of this so-called continent of the Old Red Sandstone

Distribution and divisions Devonian rocks both of marine and nonmarine origin are recognized on all continents They are thickest on sites of former geosynclines where they are commonly folded and faulted and sometimes intruded by igneous materials In the continental interiors Devonian strata are thinner and essentially horizontal Conglomerate sandstone shale and limestone are common The last two are most important Redbeds and salts are characteristic nonmarine deposits in many places

Fossils of marine organisms are abundant in the marine phase The remains of fresh water fish and land plants characterize the nonmarine phase *See* GEOSYNCLINE REDBED

Unconformities generally terminate the upper and lower boundaries In their absence transition beds are developed The Downtonian beds of England and in the Ardennes include transition sediments between the Silurian and Devonian The boundary between the Upper Devonian and Lower

nizable unconformities and the rocks of the two epochs are characterized by similar lithologies *See* UNCONFORMITY

Correlation of European and North American series and stages, and selected New York formations and groups of formations

Europe		North America (New York State)	
Stages	Characteristics at type locality	Stages	Formations
Upper Devonian		**Senecan Chautauquan**	
Fammenian (Belgium)	Shale limestone, sandy in places *Cheiloceras*, and other ammonite genera	Conewangoan Cassadagoan	Oswayo group Conneaut group
Frasnian (Belgium)	Reef limestone in certain areas *Manticoceras (Gyphyceras)* and other ammonite genera	Chemungian Finger Lakes	{Naples group {Genesee group
Middle Devonian		**Erian**	
Givetian (France)	Shale, reef limestone, *Goniatites* (ammonite) *Tentaculites Stringocephalus burtini* brachiopod fauna	Taghanican	Tully limestone
Eifelian (France)	Deep-water shales reef limestone, *Anarcestes* (ammonite)	Tioughniogan Cazenovian	Hamilton group Marcellus shale
Lower Devonian		**Ulsterian**	
Coblenzian (Germany)	Thick varied facies sandstone shale, reef limestone *Agoniatites* (ammonite)	Onesquethawan Deerparkian	Onondaga limestone Oriskany sandstone
Gedinnian (Belgium) Downtonian (England)	Sandy, clastic variegated shales Transition stage Silurian marine facies with brachiopods Old Red Sandstone facies with fresh water fishes, crustaceans land plants	Helderbergian	Helderberg group

The subdivision of the Devonian into Lower Middle, and Upper was established in Europe Later a similar nomenclature was adopted in North America The divisions are not of equal duration but are useful for description and in stratigraphic correlation European stages are based on lithologic features and on marine faunal changes chiefly those involving the ammonites (Cephalopoda) The correlation of rocks on different continents is based largely on faunal resemblances The Devonian is best known in Europe and North America because the earliest and most detailed studies were made there A correlation of the major time units and some of the more commonly known rock units is given in the accompanying table *See* CEPHALOPODA FOSSILS, STRATIGRAPHIC NOMENCLATURE

European occurrences. In western Europe two lithologic facies are developed (1) the Old Red Sandstone continental facies, where the rock strata are essentially horizontal and include interbedded marine lagoonal deposits of the Downtonian transitional stage and (2) the sandy, shaly, marine facies such as those occurring in Bohemia and Britanny where the rocks are closely folded

The Mediterranean facies are characterized by marine, deep water, geosynclinal deposits These rocks eventually were folded and metamorphosed to form Hercynian and Alpine chains The type region is Montagne Noire in France, but Devonian rocks occur throughout all southern Europe south of Britanny and the Rhenish and Bohemia massifs

In eastern Europe, Devonian rocks are transitional from the geosynclinal type of the Urals to the characteristic continental type of the Russian platform Five areas of outcrops are described (1) In the Baltic states the Old Red Sandstone facies are overlain by marine deposits of Frasnian age and these in turn by Baltic sandstones of continental origin The uppermost Devonian strata are missing (2) In central Russia, the Upper Devonian in surface outcrops consists of calcareous marine deposits with characteristic Frasnian fossils The greater number of calcareous facies than in the Baltic states implies a greater distance from the continent of Old Red Sandstone The Fammenian stage is represented by lagoonal dolomitic limestones containing faunal types that are transitional from Devonian to Carboniferous (3) The Timan outcrop consists of Middle Devonian marine strata, including the famous bituminous shales of Dormanik, and deep water facies containing the Frasnian ammonite, *Manticoceras intermescens* The uppermost beds are gypsiferous marls which contain brachiopods of Fammenian age (5) In the Urals a geosynclinal region was connected with the Mediterranean geosyncline by way of Turkestan and Asia Minor The rock series are extremely folded (Hercynian orogeny), conformable marine deposits of Lower Middle and Upper Devonian age In the Podolia area, or region of Dniester on the southwestern border of the Russian platform only Lower Devonian rocks with [illegible]

b,c) Collectively, along with the known subsurface distribution, they occupy a large portion of the continent Regionally they may be grouped as follows eastern North America Michigan Basin continental interior, and western North America

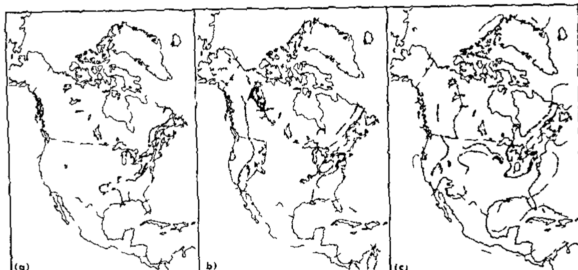


Fig 2 Paleogeographic maps of North America (a) Lower Devonian Oriskanian (Deerparkian) paleogeography (b) Middle Devonian Skaneateles Ludlowville and Moscow paleogeography (c) Middle and high Upper Devonian Chemungian Cassadagoan and Conewangoan paleogeography Shaded areas show marine water invasions Outcrops are in black (From C Schuchert Atlas of Paleogeographic Maps of North America Wiley 1955)

Eastern North America Included in this area are the Atlantic and Appalachian regions extending from the maritime provinces in Canada to Alabama in the United States Eastern New York and Pennsylvania have the thickest deposits and m st complete Devonian record Both marine and nonmarine deposits are well developed A description of the features of the New York section follows

The Lower Devonian (Ulsterian series) consists of relatively thin deposits which were limited to the axis of a geosyncline Three stages are recognized Helderbergian Deerparkian and Onesquethawan Selected formations or groups of formations for each stage and their lithologic and faunal features are as follows The Helderberg group is a primarily hard resistant limestone which forms escarpments

calcareous cement is important for glass making its thickness is variable to 300 ft locally it is limited mostly to eastern New York and is a limestone unit in Missouri The Onondaga limestone of the Onesquethawan is widespread from the Hudson Valley westward to Michigan It is 100 ft thick *thins to southwest* and is generally crystalline limestone locally shale it is fossiliferous with coral reefs widely distributed the most famous forming the Falls of the Ohio at Louisville Kentucky

The Middle Devonian (Erian series) includes widespread marine and nonmarine deposits The [illegible]

equivalent becomes calcareous somewhat petroliferous and very fossiliferous The Hamilton group a wedge of clastic detrital material 2500 ft thick in eastern New York thins westward and becomes more calcareous It is characterized by dark gray silty shale to fine sandstone and contains abundant marine fossils The upper part is a complex of red shales and conglomerate with fossils of land plants fresh water fishes and clams This is the lower part of the Catskill delta which westward merges into true marine beds (Fig 3) The middle part of the Catskill beds near Gilboa New York contains fossils of successive forest beds including trees 30–40 ft high The Tully limestone is equivalent to the Cuboides zone of Europe and is now interpreted as the homotaxial equivalent of the upper *Stringocephalus* zone of Europe

The Upper Devonian (Senecan Chautauquan series) contains very thick clastic marine and nonmarine strata resembling the Hamilton group These rocks were deposited over a longer period of time than those in the lower and middle series The New York column includes four groups of formations in ascending order Genesee Naples Conneaut and Oswayo Beds grade laterally from nonmarine redbeds in the east to black marine shales in western New York The redbeds are a continuation of the Catskill delta begun in Hamilton time and typically developed in the Catskill Mountains The greatly folded rocks crop out in long narrow bands in the folded belt of Appalachians Devonian rocks in the Acadian region of eastern Canada and extending south into Maine also are strongly folded and faulted and intruded by igneous materials

Michigan Basin This basinlike area in southern Michigan contains especially well developed Mid

dle and Upper Devonian deposits The former are dominantly calcareous and very fossiliferous the latter are chiefly shale In their faunal aspects these rocks are only partially similar to those of the New York type region The Michigan Basin was intermittently joined with the Appalachian geosyncline

Continental interior There are outcrops of Devonian rocks in widely scattered areas from Hudson Bay south to Oklahoma and Arkansas The chief outcrops are in the eastern Mississippi Valley region and Great Lakes area Deposits are dominantly limestone Although there is much black shale in Upper Devonian outcrops the beds are seldom more than a few hundred feet thick and generally rest unconformably on older rocks Numerous formations have been recognized their names and divisions vary from place to place There are abundant fossils in the calcareous phases of the Middle and Upper Devonian In the prairie provinces of Canada up to 2000 ft of salt (NaCl) and associated evaporites has been penetrated in oil drilling operations Salt deposits of Devonian age are also present at depth in Michigan

Western North America This section includes the Cordilleran province and all of northwest Canada Here the Cordilleran geosyncline extended from the Arctic Ocean to southern Arizona Sediments deposited in this former geosyncline are mostly of Middle and Upper Devonian age Limestone was dominant in the southern half of the geosyncline and shale in the northern half The thickest Devonian sections in the United States (6000 ft) are in Nevada The sediments thickened north and northwest toward the center of the former geosyncline They are especially thick in the northern part of the Canadian Cordillera where several thousand feet of Middle and Upper Devonian shale and other detrital sediments are present these are oil and gas bearing in places

Limestone layers form ramparts of the Mackenzie River There are excellent outcrops in the Canadian Rockies and a 2000 ft exposure in Mount Devon Devonian strata in the Arctic islands and in eastern Greenland may be 10 000 ft thick The earliest known land vertebrate fossils along with abundant fishes and land plants are found in Greenland deposits

Other continents Devonian strata deposits occur in central Asia east of the Urals and in China Korea southern Asia South Africa and New Zealand In the Tasman geosynclinal area in southwestern Australia there is at least 30 000 ft of Devonian sediments and igneous rocks In South America Devonian rocks are widely distributed The largest outcrop areas are in northern Paraguay and southeastern Bolivia the lower Amazon Valley and in the northern part of Colombia

Physical history (paleogeography) Devonian rocks of shallow marine origin form part of all existing continents Details of invading seas and adjoining land areas are best known in North America and Europe

North America was low and flat at the beginning of Devonian time and probably was much wider than now No mountainous tracts were inherited from the Silurian and the Canadian Shield was undoubtedly the most prominent continuous land area The region of the Cincinnati Arch along the western edge of the Appalachian geosyncline was a low land area during part of Devonian time (Fig 2) By Oriskanian times a narrow strait developed in the Appalachian trough from Newfoundland to the lower Mississipp Valley separating Appalachia from the mainland region Helderbergian age sediments were restricted to this part of the continent By Middle Devonian time a shallow sea occupied a large portion of the Mississippi Valley and extended north to Hudson Bay A seaway from the Arctic spread south into the Cordilleran geosyncline in the early part of Middle Devonian time forming a vast sea as much as 1000 miles in width This sea reached eastward to Manitoba Michigan Illinois and Iowa by Marcellus time bringing in the *Stringocephalus* fauna and other organisms characteristic of European and Asiatic upper Middle Devonian The eastern and western geosynclines may have been connected at this time An estimated 40% of the continent was inundated In late Middle Devonian an uplift known as the Acadian disturbance began in the eastern borderlands This uplift formed a continu

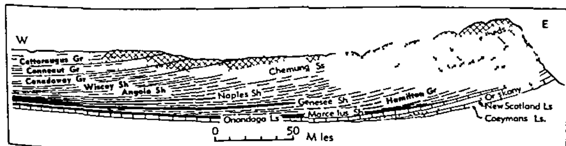


Fig 3 Diagrammatic section of Catskill delta Continental redbeds interfinger with nearshore marine deposits and these in turn grade into offshore sediments Progressive westward shift of these enviro during Devonian time is indicated by the upward lateral displacement of the different facies (*From R C Moore Introduction to Historical Geology 2d*

ous mountain chain from Acadia to Cape Hatteras by the end of the Devonian period Sediments eroded from its western slopes to form the Catskill delta A large delta formed contemporaneously in the Gaspe region Igneous activity accompanied the uplift in Acadia and New England The Acadian disturbance terminated the major sea invasions in the Appalachian geosyncline

Physical events in other parts of the continent are not as well known In widespread areas in northwest Canada most Upper Devonian rocks have been removed by erosion The remainder are over lain unconformably by Cretaceous rocks Thicken ing of Devonian sediments westward and their clas tic character indicate highland areas west of the present continental border Sedimentation probably was continuous into Mississippian time in those areas where lithologic and faunal changes cannot be differentiated Gradual withdrawal of shallow seas from the remainder of the continent and emer gence of land followed the events of Devonian time

Climate The Devonian climate was variable but generally mild and without strongly marked cli matic belts Proof of this is (1) the abundance of life and distribution of similar plant and animal species regardless of latitude (2) second greatest coral reef development in geologic history (3) Arctic migration of *Stringocephalus* fauna and (4) similarity in land plants traced from the British Isles and Spitsbergen in Europe by way of east Greenland to eastern New York

Redbeds of the Catskill delta the Old Red Sand stone and contemporaneous Devonian redbeds in other parts of the world developed in a mild humid climate with seasonal rainfall This dry climate probably accounts for the great thickness of salts deposited in highly saline restricted seas *See* EVAP ORITE (SALINE) PALEOBOTANY PALEONTOLOGY [C A S]

[illegible] th [illegible] Ar [illegible] 19[illegible] *The Old Red Sandstone* reprint 1959

Dew

Drops or films of water formed by condensation of water on outside exposed surfaces especially of vegetation Hoar frost is the corresponding phenom enon at temperatures below freezing In most cases not enough water is collected to be recorded as pre cipitation But at certain locations in Palestine the annual total has been estimated at around 200 mm (8 in) enough to permit growth of summer crops that would otherwise require irrigation

Dew forms on clear nights when there is cooling by radiation provided exposed surfaces cool below the dew point of the air Moisture then condenses on the exposed surfaces If the ground is wet mois ture is evaporated which adds to the supply and increases the deposit of dew on vegetation Fog also deposits moisture which unlike dew collects inside trees and similar places as readily as on surfaces exposed to the sky Deposits similar to those pro duced by fog may form when cold air is replaced suddenly by warmer and more moist air *See* DEW POINT HUMIDITY PRECIPITATION (METEOROLOGY) RADIATION TERRESTRIAL, VAPOR PRESSURE [J R F]

Bibliography R Geiger (tr M M Stewart) *The Climate Near the Ground* 1950, H Landsberg *Physical Climatology* 1941

Dew point

The temperature at which air becomes saturated when cooled without addition of moisture or change of pressure Any further cooling causes con densation Fog and dew are formed in this way

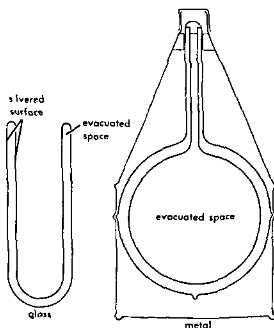
Frost point is the corresponding temperature of saturation with respect to ice At temperatures be low freezing both frost point and dew point may be defined because water is often liquid (especially in clouds) at temperatures well below freezing at freezing (more exactly at the triple point + 01°C) they are the same but below freezing the frost point is higher For example if the dew point is −9°C the frost point is −8°C Both dew point and frost point are single valued functions of vapor pressure

Determination of dew point (or frost point) can be made directly by cooling a flat polished metal surface until it becomes clouded with a film of wa ter or ice the dew point is the temperature at which the film appears In practice the dew point is usually computed from simultaneous readings of wet and dry bulb thermometers *See* DEW, EVAPO RATION FOG HUMIDITY VAPOR PRESSURE [J R F]

Bibliography R J List (ed) *Smithsonian Me teorological Tables* 6th ed rev 1951

Dewar flask

A vessel having double walls the space between being evacuated and the surfaces facing the vacuum being heat reflective It was invented in 1892 by



Typical Dewar containers

Sir James Dewar as a container for liquid oxygen

Dewar's original flasks were made of glass with a coating of mirror silver this type is still used in the laboratory But for shipment and storage of liquid gases metal vacuum vessels are used (see illustration) Metal vessels with a capacity of 50 liters can preserve liquid oxygen with an evaporation loss of only 4% per day Evaporation rates in 110 000-liter vessels designed for transport by rail way are approximately 0 1% per day for oxygen and 0 8% per day for hydrogen

Thermos Bottle is a trade-mark for a Dewar vessel for hot and cold foods [H W R]

Dewaxing (petroleum)

The process used in petroleum refining to separate those hydrocarbons that readily solidify (waxes) from petroleum These processes give a wax and a so-called wax free fraction The wax free portion always has a reduced or lower temperature of fluidity or pour point than the original fraction thereby giving improved flow characteristics

The removal of wax from petroleum fractions is one of the most important steps in the production of lubricating oils and fuel oils of low pour point Lubricating oils of low pour are needed in all types of industrial lubrication and power propulsion equipment Without a lubricant that will flow readily at low temperatures to the various points of application modern industry would be handicapped

At the present time, the following processes are currently being used in the dewaxing of petroleum cold pressing centrifuge dewaxing solvent dewaxing and complex dewaxing The principal characteristics of these processes are briefly given in the table *See* PETROLEUM PROCESSING; WAX PETROLEUM

[W E K]

Bibliography American Chemical Society (Division of Petroleum Chemistry) *Progress in Petroleum Technology* Advances in Chem Ser 5 1951

Dextran

A polysaccharide synthesized by bacteria including those belonging to the genera of *Leuconostoc*, *Streptococcus Acetobacter* or related forms (*see* LACTOBACILLACEAE) Its principal utility now lies in its ability to serve as a blood plasma volume expander Formerly high molecular weight dextran was regarded as a nuisance around sugar refineries Dextran's chemical and physical properties depend upon both the strain of microorganism employed and the environmental conditions imposed upon the bacterium during growth or the reaction conditions where an enzymatic method of dextran production is employed *Leuconostoc* and *Streptococcus* species convert sucrose to dextran and fructose primarily *Acetobacter* species convert dextrin to dextran the α 1 4 linkage is converted to an α 1 6 linkage *See* POLYSACCHARIDE

Commercial petroleum dewaxing processes

	Cold pressing	Centrifuge dewaxing	Solvent dewaxing	Complex dewaxing
Petroleum stocks to which applicable	Low viscosity pressable distillates	Residual stocks or bright stocks	Wax distillates residual stocks	Low viscosity distillates*
Solvents used	None	Petroleum naphtha	Acetone-benzene-toluene methyl ethyl ketone-benzene-toluene liquid propane benzene-ethylene dichloride trichloroethylene	Methanol or a suitable complexing solvent for urea (water necessary as a component for some solvents)
Means of separation	Pressure filters	Centrifuge or precoated filters (precoat or filter aid)	Pressure filters or vacuum filters	Wax forms a urea adduct mixture which is removed by filtration
Temperature range	+30 to 0°F	+10 to −50°F direct expansion ammonia chillers used with exchangers	+50 to −40°F depends on dewaxing differential and pour desired	No refrigeration necessary—operates at essentially room temperature
Comments	Wax cake from this operation used as charge in sweating process to make petroleum wax	Process also used for dewaxing and deoiling of wax or petrolatum	Crystal modifiers may be used to accelerate wax crystal formation and filtering rate selected slack waxes used in sweating or solvent fractionation processes this process also used for deoiling of petrolatum	Process has limited applicability

* Refrigerator oils transformer oils hydraulic oils.

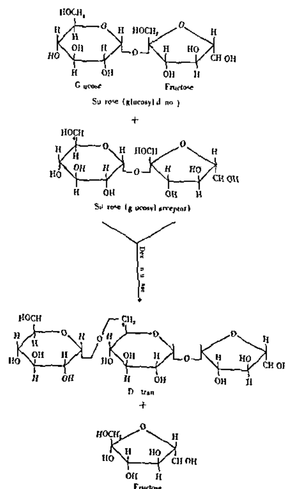
The conversion of sucrose to dextran and its by product fructose is a transglucosylation reaction This transformation is mediated by an enzyme dextran sucrase It is readily obtained extracellularly from suitable strains of *Leuconostoc mesenteroides* propagated under appropriate conditions *L mesenteroides* NRRL B 512F has been adopted by nearly all of the Western nations for the production of dextran

Dextran is similar to glycogen and amylopectin in that it is a branched glucose polymer However it differs from amylopectin in that the principal

[illegible]

NRRL B 512F forms a polymer with about 95% α-1 6 type of linkage The ratio between the 1 6 and the non 1 6 linkages may vary from 20 to 1 to almost 1 to 1 depending upon the bacterial strain and is characteristic of the enzyme obtained from it

Mechanism Formation of dextran from sucrose involves the following reaction



The reaction requires two substrates Dextransucrase is more specific in its *requirement* for the glucosyl donor substrate (sucrose) than for its acceptor substrates The acceptor substrate may be sucrose maltose isomaltose α methylglucoside and low molecular weight dextran The average molecular weight of the dextran polymerized is controlled by both the quality and quantity of the acceptor substrate incorporated either in the fermentation medium or in the enzymatic reaction

The kinetic constants can be varied at will since dextran is formed from a double substrate system (*see* ENZYME) Neither the V_{max} nor the Michaelis constant for the polymerization reaction is fixed, they are varied by alteration in the concentration of either cosubstrate (glucosyl donor and glucosyl acceptor) In some respects the mechanism of the polymerization is similar to that of the condensation type in that water is eliminated It also displays however certain characteristics of the chain reaction type in that the molecular weight of the dextran produced is essentially that of the finished product

Production process The average molecular weight of the dextran produced by most of these organisms is generally on the order of several to hundreds of millions One strain of *Streptococcus* produces however a dextran with a molecular weight of about 100 000 or slightly lower The high molecular weight dextran as ordinarily produced by *L mesenteroides* is hydrolyzed to a product with an average molecular weight of about 75 000 for use as a blood plasma volume expander The heads and tails must be discarded More information is available on the utility of dextran from *L mesenteroides* NRRL B 512F for use as blood plasma substitute than on the dextran from the *Streptococcus* spp For use as plasma extender the high molecular weight dextran is acid hydrolyzed fractionated (as to its molecular weight distribution) and purified The specifications on the dextran for use as a plasma volume extender vary in different countries, the physical chemical and biological specifications it must meet are extremely rigid *See* INDUSTRIAL MICROBIOLOGY POLYMERIZATION [H M T]

Dextrin

A polymer of D glucose which is intermediate in complexity between starch and maltose The dextrins are usually obtained by hydrolysis of starch with diastase (amylases) The higher dextrins resemble starch while the lower dextrins more nearly resemble the sugars Compared with the original starch the dextrins produce less viscous solutions They are soluble in water but insoluble in alcohol Dextrins may be obtained from starch by controlled hydrolysis with acids The Lintner method for solubilizing starch consists of subjecting the native starch grains to 7 5% hydrochloric acid at room temperature for 7 days The degradation product (dextrins) thus produced readily dissolves in water, but still gives a blue color with iodine More drastic treatment of starch with acid will produce dextrins having a purple red or no color with iodine Dextrins are used commercially

as adhesives Tapioca waxy maize and sweet po tato starch represent the best material for their manufacture *See* GLUCOSE

When a starch is exposed to the action of *Bacil lus macerans* or to the bacteria free filtrate of this microorganism a mixture of water soluble dex trins known as Schardinger dextrins is produced From this mixture three distinct nonreducing crystalline compounds can be isolated These dex trins α β and γ are known to possess cyclic structures consisting of six seven and eight 1 4-α glucosidically linked D glucose units respec tively *See* CARBOHYDRATE [W Z H]

Diabase

A fine-textured dark gray to black igneous rock composed mostly of plagioclase feldspar (labra dorite) and pyroxene and exhibiting ophitic tex ture It is commonly used for crushed stone Its resistance to weathering and its general appearance make it a first class material for monuments

The most diagnostic feature is the ophitic tex ture in which small rectangular plagioclase crys tals are enclosed or partially wrapped by large crystals of pyroxene As the quantity of pyroxene decreases the mineral becomes more interstitial to feldspar The rock is closely allied chemically and mineralogically with basalt and gabbro As grain size increases the rock passes into gabbro as it decreases diabase passes into basalt This intermediate characteristic justifies classifying dia base as a hypabyssal rock

Diabase forms by relatively rapid crystallization of basaltic magma (rock melt) It is a common and extremely widespread rock type It forms dikes sills sheets and other small intrusive bodies The Palisades of the Hudson near New York City are formed of a thick horizontal sheet of diabase In the lower part of this sheet there are rich con centrations of olivine and pyroxene commonly be lieved to have formed as heavy crystals settled through the molten diabase

As defined diabase is equivalent to the British term dolerite The British term diabase is an al tered diabase in the sense defined here *See* BASALT GABBRO IGNEOUS ROCKS

[C A C]

Diabetes

A term which indicates excess excretion of some body substance but which is commonly used to in dicate diabetes mellitus a metabolic disorder aris ing from a defect in carbohydrate utilization by the body related to an inadequacy or abnormality of insulin production by the pancreas Many other factors are involved including both pituitary and adrenal hormone regulation and the full explana tion for diabetes is not available Therapeutic measures however have advanced to the point where most diabetic patients can expect a nearly normal life span and productivity *See* CARBOHY HYDRATE CARBOHYDRATE METABOLISM HORMONE INSULIN

Diabetes mellitus classically is marked by ex cessive urinary output thirst and hunger In creased susceptibility to infections weight loss constitutional symptoms and a higher incidence of other diseases are common in diabetics In ad vanced cases complications involving the heart blood vessels kidneys and liver may produce seri ous illness or death

The pathologic lesions may be nonspecific but advanced cases usually show characteristic changes in the organs mentioned Clinically blood sugar levels are elevated there is sugar in the urine and glucose tolerance tests are abnormal *See* CLINICAL PATHOLOGY

Several subtypes of diabetes mellitus are recog nized particularly the juvenile form the senile form and the diabetes associated with obesity

Diabetes insipidus is unrelated to diabetes melli tus It results from a hypothalamic pituitary defect which causes large amounts of water to be lost in the urine It is uncommon and may follow trauma and tumors that cause local brain damage

Amino acid diabetes is a rare congenital defect of kidney tubule reabsorption in which there is ex cessive excretion of certain amino acids and other substances in the urine *See* AMINO ACIDS

[E G S T]

Diadematacea

A superorder of Euechinoidea having a rigid or flexible test perforate tubercles aulodont lantern complete perignathic girdle and branchial slits (*see* ECHINOIDEA) The group arose in the Late Triassic and differentiated into three stocks the Diadematoida Echinothurioida and Pygasteroida In the two former the anus remained within the apical system in the latter the anus migrated into the posterior interambulacrum *See* IRREGULARIA

[H B F]

Diadematoida

An order of Diadematacea with hollow primary radioles and diademoid ambulacral plates (*see* ECHINOIDEA) The tubercles are normally crenulate and the anus remains within the apical system Three families comprise this order The Diademat idae with crenulate tubercles are mainly large long spined purple black echinoids which inhabit tropical and subtropical seas They extend from the Jurassic to the present day *Diadema* is an exam ple The Micropygidae includes only the deep wa ter Indian Ocean genus *Micropyga* which has noncrenulate tubercles and umbrellalike o iter tube-feet The Aspidodiadematidae have crenulate tu bercles and the axial tube of the radioles are locu late They are mainly Jurassic and Cretaceous but survive as deep sea forms *See* DIADEMATACEA

[H B F]

Diagenesis

Those processes that alter the structure texture and mineralogy of a sediment during its depo tion lithification and ultimate burial but wh

exclude high temperature and high pressure modifications attributed to metamorphism Diagenetic changes are intergradational with those of metamorphism at elevated temperatures and pressures and with those of weathering at or near surface conditions Diagenetic processes tend to establish chemical equilibria between minerals in the accumulating sediment the incorporated biota and its shells and interstitial fluids during three somewhat separate stages in the history of a sedimentary rock

Stages of diagenesis The stages of diagenesis can be regarded as threefold (1) In the initial or depositional modification in the raw detritus is controlled by the environment of the water sediment interface (2) The intermediate or early burial stage is confined to changes which occur in the upper few feet of accumulating debris This stage is transitional with the condition of lithification as well as the initial stage but represents a time of major modification in bedding texture and to a lesser degree in mineralogy (3) The late burial or premetamorphic stage is best developed by the environment of deep burial where temperatures above the boiling point of water and pressures of thousands of pounds per square inch promote important changes in bedding cleavage porosity cementation and mineral authigenesis following initial compaction (*see* AUTHIGENIC MINERALS)

Initial stage An important control of the environment of the initial stage is the salinity of the fluid although local chemical changes are influenced by the bottom dwelling organisms depth of water and intensity of currents The condition is that of an open system chemical reactions between the sediment and water do not attain equilibrium and there is active solution of unstable minerals Reactions proceed toward simplification of products namely reduction in the variety of minerals in the raw detritus and concentration of *resistates* (quartz) precipitates (carbonates such as the contribution of epifauna) and hydrolyzates (clays)

Early burial The environment of early burial is a system which is partially closed and important chemical conditions are introduced by sediment dwelling organisms and interstitial *fluids* *The* latter are not [illegible]

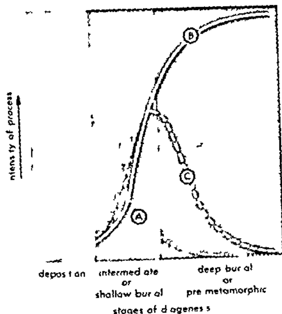
Late stage The late stage of diagenesis is reflected by the nature of equilibria brought about by depth of burial Shallow burial produces virtually no change in grain to grain contacts after lithification whereas deep burial produces intergranular penetration and a fabric associated with metamorphic rocks Sandstones of shallow burial tend to become aquifers and originally saline pore waters are gradually diluted This process promotes intrastratal solution indicated principally by pitting frosting and rounding of mineral grains Some sediments remain at great depths for long periods of time and intrastratal solutions of high ion concentration react with detrital minerals to develop new minerals by authigenesis Such minerals provide the basis for the concept of diagenetic grades representing a measure of the progress of diagenesis in the development of textures and mineralogy each of which reflects increasingly high temperatures pressures and ion concentrations

Diagenesis in sandstones Among quartzose sandstones the initial stage is characterized by rounding and solution pitting of quartz Chert is unstable and is reduced in amount Early burial is characterized by precipitation of overgrowths on loose quartz grains but actual cementation or interlocking of grains is not an important process Commonly precipitation of small amounts of carbonate as cement follows precipitation of quartz overgrowths Where such sandstone passes laterally into limestone increasing deposition of carbonate locally forces quartz grains apart and produces a texture recognized as "floating" sand grains The late stage of diagenesis associated with deep burial is recognized either by strong addition of carbonate cement and concomitant solution of quartz grains or in the absence of carbonate cement intergranular penetration of grains to produce sutured boundaries (*see* STYLOLITES)

Argillaceous sandstones display a commonplace diagenetic sequence During early burial silica is deposited on some of the large grains as overgrowths Chert becomes metastable and tends to precipitate locally Siderite concretions are considered to develop during this stage in brackish water or marine sediments where the local pore water contains a high concentration of carbonate ions Deep burial is indicated principally by introduction of carbonates corroding quartz and replacing the clay matrix Chertification of the matrix detritus principally clay is typical of the intermediate and late stages of diagenesis Also certain clay minerals become unstable and tend to recrystallize into larger flakes which can be identified as chlorite muscovite and biotite

Typical graywacke sandstones show well developed reaction effects between mineral grains and destruction of original boundaries Concentration of authigenic chlorite suggests that most graywackes have attained the chlorite grade of metamorphism The principal diagenetic relationships such as welding of grains and recrystallization of clays are considered to be late in development and attendant upon deep burial and elevated temperature

Diagenesis in shales Diagenesis in shales is manifested primarily by the extent of separation along the bedding recrystallization of clay minerals and development of micas of restricted composition Parting along bedding is accentuated by carbonization of organic debris particularly woody material and is inhibited by precipitation of carbonate and silica Although fissility is inherited from the detritus and the nature of the deposition



Diagrammatic representation of intensity of major diagenetic processes Curve A rounding pitting and frosting of grains overgrowths on quartz solution of detrital chert reactions controlled by Eh and pH Curve B grain interpenetrations mineral authigenesis carbonate cementation recrystallization development of fissile bedding Curve C, development of concretions compaction silicification and dolomitization

it is accentuated by loading and unloading of strata and by orientation of clay minerals and organic remains during compaction As long as the rock is under significant load bedding plane cleavage is poorly developed but upon unloading and during weathering the incipient planes of failure become enlarged until parting occurs Also such parting is gradational with fracture cleavage inasmuch as not all of the separation planes are coincident with bedding

Compaction is an important diagenetic process in all shales particularly those which are deposited as hydrosols Loading of such material causes destruction of the sol and ejection of water until minerals of silt and clay dimension are brought into contact, after which reduction in pore water with time and loading approaches the zero limit asymptotically

Early burial is marked by reactions controlled by pH (alkalinity acidity) and Eh (oxidation reduction potential) Dark gray colors and sulfide minerals signify the existence of reducing conditions whereas light gray and red colors and oxides and hydroxide minerals are associated with oxidizing environments Calcite is the principal indicator of the pH (precipitated at approximately 7 8)

Principal modifications during late burial are alteration of clay minerals to micas and recrystallization to produce a welded aggregate of crystals in anisotropic orientation Shales associated with volcanic rocks often are very siliceous the silica having been precipitated during late burial and believed to have been derived from waters moving through the volcanic sequence

Diagenesis in carbonates Diagenesis in carbonates is dominated by recrystallization and wholesale introduction of magnesia and silica Recrystallization of fossil fragmental texture destroys original organic structures and substitutes an interlocking crystal meshwork held without cement The course of recrystallization is progressive from stages in which fossil fragments are attached by a crystalline cement to complete alteration to a granoblastic texture

Selective introduction of magnesia to form dolomite throughout entire formations is known in the majority of cases to postdate precipitation of siliceous concretions Dolomitization is developed preferentially in sites of broad structural arches which were tectonically stable for long periods

Introduction of silica is principally in the form of chert as nodules and lenticular beds generally emplaced in stratigraphically favorable positions Mineral paragenetic relationships favor the interpretation that such silica is introduced during early burial and preceding lithification

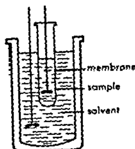
See SEDIMENTARY ROCK SEDIMENTATION (GEOLOGY) [E C D]

Bibliography F J Pettijohn *Sedimentary Rocks* 2d ed, 1957, Z L Sujkowski Diagenesis *Bull Am Assoc Petrol Geologists* 42(11) 2692 2717 1958

Dialysis

A process by which dissolved substances can be separated from colloids by diffusion through a membrane The dialyzing membrane acts as a sieve depending upon the size of its pores and also functions because of differences between the diffusion and adsorption of the large and small particles

The simplest dialyzer consists of a sheet of parchment paper regenerated cellulose or collodion which is placed over the open end of a tube containing the mixture of colloid and solute The tube is then inserted in a bath of pure solvent The smaller dissolved ions and molecules can pass through the membrane leaving the colloidal particles inside the tube Thin sacks of collodion have been used to provide a larger surface for transfer



for dialysis

The rate of dialysis depends upon the area of the dialyzer the size of the pores the temperature the electric charges and the relative concentrations of the solutions on the two sides of the membrane

In electrodialysis the dialyzing chamber is placed between two electrodes with pure water in compartments on either side Under the influence of a direct current the charged ions migrate from the sample solution to the oppositely charged electrodes *See* COLLOID DONNAN EQUILIBRIUM

Diamagnetism

That branch of magnetism which treats of diamagnetic phenomena and of the properties of diamagnetic bodies Diamagnetism is a property exhibited by substances with a negative magnetic susceptibility that is by substances which magnetize in a direction opposite to that of an applied magnetic field (*see* SUSCEPTIBILITY MAGNETIC) A diamagnetic substance has a magnetic permeability less than 1 and is repelled when placed near a magnet The magnetization of diamagnetic substances is associated with the currents induced on application of a magnetic field According to Lenz s law the flow of an induced current is in such a direction as to oppose the change of flux of the inducing field this accounts for the negative susceptibility The diamagnetic susceptibility is invariably small of the order of -10^{-5} cm³/mole *See* LENZ S LAW PERMEABILITY MAGNETIC

All matter responds to applied fields in this diamagnetic fashion However some substances also have net electronic orbital or spin magnetic moments or both which can be aligned by an applied magnetic field in a direction along (not opposite to) the field this property is called *paramagnetism* (*see* PARAMAGNETISM) For these substances the observed susceptibility χ is the sum of diamagnetic and paramagnetic terms

$$\chi = \chi_d + \chi_p \qquad (1)$$

Under ordinary conditions χ_d is temperature independent Hence if χ_p follows the inverse temperature dependence of Curie s law one can experimentally determine the separate *contributions of* χ_d and χ_p by measuring χ as a function of temperature However if χ_p is also temperature independ

[illegible]

[illegible] are the alkali and alkaline earth metals where χ_p is unusually small and of the order of χ_d and salts or solutions containing only a small fraction of paramagnetic atoms The condition for $\chi_p = 0$ *and hence* pure diamagnetism is that all electron [illegible]

electrons satisfy this condition, an important exception is O_2 The condition is also satisfied by most nonmetallic solids except compounds containing atoms with incomplete inner shell electron groups such as the transition rare-earth and actinide elements *See* ELECTRON SPIN

As stated previously the diamagnetic response of a substance is small, only a very small fraction of the applied magnetic field is shielded from the interior of the substance by the induced diamagnetic currents There is one case however, in which the inducing field is completely shielded (except for small surface effects) This is the perfect diamagnetism exhibited by superconductors and is known as the Meissner effect (see MEISSNER EFFECT SUPERCONDUCTIVITY)

Langevin theory The Langevin theory of diamagnetism (P Langevin 1905) is based on the idea of inducing an electronic current inside an atom The theory employs the Larmor theorem (*see* LARMOR PRECESSION) In a magnetic field H, the precession of the Z electrons within the atom is equivalent to a current equal to $-Z(e/c)(\omega_L/2\pi)$ in electromagnetic units Here e c is the magnitude of the electronic charge in emu and ω_L is the angular Larmor frequency

$$\omega_L = -eH/2mc \qquad (2)$$

where m is the electronic mass The magnetic moment μ arising from this induced current is equal to the product of the current and the area of the *current loop or*

$$\mu = -Z(e/c)(\omega_L/2\pi)(\overline{\rho^2}) \qquad (3)$$

where $\overline{\rho^2}$ is the statistical average over a large number of atoms of the square of the perpendicular distance of an electron from the field axis This average is equivalent to $\overline{x^2} + \overline{y^2}$ if H is along z For a random assembly of atoms since $\overline{x^2} = \overline{y^2} = \overline{z^2}$ one may write

$$\overline{\rho^2} = (2/3)(\overline{x^2} + \overline{y^2} + \overline{z^2}) = (2/3)\overline{r^2}$$

where $\overline{r^2}$ is the mean square distance of the electron from the nucleus Thus the diamagnetic susceptibility of N atoms is given by

$$\chi_d = N\mu/H$$
$$= -(Ze^2N/6mc^2)\overline{r^2} \qquad (4)$$

This is P Langevin s result as corrected by W Pauli The molar susceptibility χ_M is obtained by replacing N in Eq (4) by Avogadro's number Numerically

$$\chi_M = (-2\,83 \times 10^{10}\ \text{cm/mole})\Sigma\overline{r^2} \qquad (5)$$

where the summation is to be taken over all the electron orbits in the atom Since $\overline{r^2}$ is of the order of 10^{-15} cm² this gives $\chi_M \sim -10^{-5}$ cm³/mole

Langevin s formula is not modified by quantum mechanics and the problem becomes that of determining $\overline{r^2}$ of the electronic wave function The calculation for many electron atoms is quite in

volved and experimental values of $\overline{r^2}$ as determined by using Eq (4) or (5) give a very useful check of the nature of the wave function for large r This complements x ray and electron diffraction data which give information for the most part only for small r

Ionic crystals In the case of diamagnetism in ionic crystals the Larmor theorem holds for the individual ions The diamagnetic susceptibility may be computed with reasonable accuracy from the sum of the individual ion susceptibilities

Diamagnetism in rare gases and in rare gas configurations of ions in ionic crystals is shown in the accompanying table These are measured

Molar diamagnetic susceptibilities (all $\times 10^{-6}$ cm³/mole)

He	−1 9	Li	−0 7	Mg	−4 3	F	−9 4
Ne	−7 2	Na	−6 1	Ca	−10 7	Cl	24 2
Ar	19 4	K	−14 6	Sr	−18 0	Br	−34 5
Kr	−28	Rb	2[illegible] 0	Ba	−29 0	I	−50 6
Xe	−43	Cs	−35 0				

values calculations by D R Hartree E C Stoner J C Slater and others are in reasonable agreement

Molecules In most molecules the electrons are not moving in a single field of force and the Lar

the presence of a mean square magnetic moment although the mean moment vanishes Calculations for more complicated molecules are exceedingly difficult In aromatic ring molecules such as benzene with H normal to the ring the electrons can precess around the ring or at least in partial ring like orbits about many nuclei This gives rise to a much larger susceptibility than is possible when H is parallel to the ring Crystals with layerlike structures such as antimony bismuth and graphite also exhibit large anisotropies in diamagnetic susceptibilities

Bohr–van Leeuwen theorem This theorem (N Bohr 1911 J H van Leeuwen 1919) proves the complete absence of magnetism in classical theory For bound electrons this comes from a cancellation of paramagnetic and diamagnetic susceptibilities providing one does not (as did Langevin) implicitly quantize the paramagnetic moments by setting them all equal to μ For a system of free electrons confined to a box the induced diamagnetic currents in the interior of the box are just cancelled by the currents from electrons which bounce in cuspoidal paths off the walls Thus magnetism is inexplicable in classical physics and is a quantum phenomenon

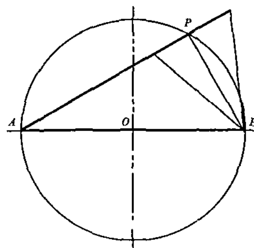
Free electrons The diamagnetism of free electrons which vanishes classically has been calculated quantum mechanically by L Landau (1930) For particles obeying Fermi Dirac statistics such as the electrons in a metal the numerical value of the Landau diamagnetism is exactly one third of the Pauli spin paramagnetism

Bound electrons The diamagnetism of bound electrons in other than ionic crystals is difficult to calculate In metals the diamagnetism is a sum of contributions from the nonconducting core electrons for which the Langevin theory is adequate and from the conduction electrons for which the Landau theory must be modified to take account of the periodic potential from the ion cores Metals in the bismuth group have unusually high diamagnetic susceptibilities ($\sim -10^{-4}$ cm³/mole) coming from the conduction electrons These metals also show the de Haas–van Alphen effect which is a quasi periodic variation of susceptibility when plotted against $1/H$ at low temperatures The susceptibility may even oscillate between diamagnetism and paramagnetism [E A F K]

Bibliography C Kittel *Introduction to Solid State Physics* 2d ed 1956 J H Van Vleck *The Theory of Electric and Magnetic Susceptibilities* 1932 A H Wilson *The Theory of Metals* 2d ed 1953

Diameter

Any chord of a circle that contains the center of the circle It is also a chord of maximum length (This property furnishes a basis for extending the concept to more general figures for example the diameter of any point set is the least upper bound



Circle of diameter AB

of the distances of two points of the set) If A B are the end points of a diameter of a circle then angle APB is acute right or obtuse ($A \neq P \neq B$) according as P is outside on or inside the circle respectively The length of a circle (circumference) is πd where d = length of diameter and the area inclosed is $\pi d^2/4$ *See* CIRCLE [L M BL]

Diamond

A mineral composed entirely of the element carbon crystallized in the isometric system

Physical properties Gem diamonds have a d sity of 3 53 but the tough black cokelike

gates of microscopic crystals sold as carbons and known to the layman as black diamond may have a density as low as 3 15 It is the hardest known sub stance Boron nitride which has been synthesized by the General Electric Company and called bora zon is the second hardest material known Boron carbide silicon carbide tungsten carbide and aluminum oxide rank below boron nitride in hard ness in that order *See* CEM MINERALOGY

Because these substances are all crystalline the bonds between atoms are arranged in definite pat terns thus certain planes and directions across a crystal surface have greater concentrations of bonds than others Therefore hardness varies with the direction of abrasion Diamond crystals can be cut only by diamond dust on a lap or rapidly rotat ing horizontal plate when the softer directions of the diamond crystal are presented to the diamond particles that attack it In the random distribution of diamond dust on a lap some will present their hardest directions to the diamond being cut It is said by the General Electric Company that the hardest directions of borazon are harder than the softest directions of diamond

At a temperature of about 900°C in an oxygen atmosphere diamond slowly burns to carbon diox ide At 1000°C in an inert atmosphere it inverts slowly to graphite and at 1700–1800°C the rate of inversion is very rapid Borazon is greatly superior to diamond in heat resistance

All except a few diamonds are nonconductors of electricity but all are excellent heat conductors superior to iron and steel Some diamonds will con duct electricity only when subjected to radioactive emanations The current conducted is proportional to the intensity of radiation Devices (proportional counters) using this principle have been developed for measuring radioactivity Under intense radio active bomb [illegible]
then brown [illegible]
the process [illegible]
heat [illegible]

Crystal structure Diamond crystallizes as octa hedrons dodecahedrons and cubes the first two forms being by far the most common Overgrowths of dodecahedrons and cubes on octahedrons are not uncommon Some crystals from Sierra Leone and the Belgian Congo show all three forms nearly equally developed These are invariably opaque yellow to brown crystals often showing a concen tric layering with clear colorless octahedral dia mond in the center In the trade they are known as coated stones The outer layers are slightly im pure diamond which is usually much twinned

Some evidence of twinning parallel to the octa hedron faces is present in nearly every crystal The crystals of highest purity are often irregularly shaped Those that are faintly yellow in color have the [illegible]
ofte [illegible]
is o [illegible]
of [illegible] foreign matter generally called carbon spots Twinning is so common that diamond cutters refer to the most obvious as macles Large twinned areas are called blocks and the smaller ones knots or pings

There is an extensive literature on type I and type II diamonds Two positions have been taken by advocates of this classification (1) that they represent two distinct types and (2) that there are only degrees of type II ness in some crystals They can definitely be differentiated by absorption dif ferences in the ultraviolet or the infrared Some type II diamonds are also semiconductors It has been shown that nitrogen atoms may randomly sub stitute for carbon in the diamond structure and that the degree of absorption in the infrared can be correlated with the amount of nitrogen substitu tion The much more common type I diamonds show this substitution of nitrogen in the crystal structure while in the strongly type II stones it is not found This evidence supports the theory that type II ness is a matter of degree English and In dian scientists are the strongest advocates of type II diamonds as distinct entities

It is now generally accepted that diamond crys tals have an octahedral structure (*m3m*) For many years the idea of a tetrahedral structure persisted although it was contrary to x ray and etch figure evidence The tetrahedral theory developed from the external morphology of diamonds from the Kimberley district by acceptance of the unusual rather than the commonplace characteristics

Octahedral crystals are typical of the better qualities of diamond from Sierra Leone Ghana Angola and the Belgian Congo Dodecahedral crystals are typical of Brazil Octahedrons and dodecahedrons with the latter predominating are typical of the Kimberley district Irregular shapes are characteristic of South West Africa and Tan ganyika although these usually reveal good exter nal evidence of crystal forms

Occurrence Primary diamonds are invariably found in a typical basic high magnesia igneous rock to which the name kimberlite has been given Kimberlite is found in dikes and pipes Productive dikes have been mined in Sierra Leone the Orange Free State and the Transvaal Pipes the necks of old volcanoes which are roughly elliptical in plan have been mined in the Kimberley district of Cape Province and Orange Free State in the Transvaal at the Premier Mine in Tanganyika at Mwadui and Mabuki Shinyanga District in the Belgian Congo at Bakwanga in Sierra Leone near Yen gema and in the United States at Murfreesboro Arkansas

Several hundred dikes and pipes of kimberlite that contained diamonds have been found Only a few have been profitable to mine at depth All pipes have an enriched surface layer which has been de veloped by erosional processes Initially the upper few feet of many pipes may be profitably mined although they may not warrant mining at depth The [illegible]
ar [illegible]

and Angola conglomerates of an early geological age have been profitably mined These are the exceptions and most alluvial diamonds are recovered from modern gravels All of these are stream gravels except the beach gravels of the Atlantic Ocean extending from Conception Bay on the north to Buffels River on the south in South West Africa and Namaqualand In this area the deposits adjacent to both sides of the Orange River mouth are the richest *The beaches that have been mined are from a* few feet to 500 ft above the present sea level The productive gravels are usually covered with tens of feet of drifting dune sand

History Diamonds were mined from stream gravels in India and Borneo in prehistoric times Originally the words adamas adamant and diamant were given to the very hard colorless transparent minerals now known as diamond corundum spinel topaz, and quartz. Pliny describes the geometrical shape of six varieties of adamas one of which is obviously the mineral now known as diamond The authentic history of diamond mining begins with Jean Baptiste Tavernier s visit to Golconda India (1638–1668)

About 1720 diamonds were identified in the gold washings of the Jequitinhonha River near Diamantia Minas Gerais about 300 miles due north of Rio de Janeiro Brazil For a century and a half this district and the area near the headwaters of the Paraguay River in the State of Mato Grosso Brazil were the chief sources of the world s diamond supply Diamonds have been found in every state in Brazil and along the northward flowing tributaries of the Orinoco River in Venezuela and British Guiana Dodecahedral crystals are characteristic *of Brazil* and the name *Brazilian diamonds* as now used describes this shape and not the geographic origin of the diamonds thus designated All the Brazilian production is from placers either ancient or modern

In 1866 the first South African diamond was identified It was a 21¼-carat stone among the playthings of a small boy living near the banks of the Orange River at Hopetown (*see* CARAT) In 1868 a small diamond was found 80 miles north of Hopetown at the German mission of Pneill on the banks of the Vaal River a tributary of the Orange The village of Klipdrift now known as Barkly West across the river has since become a center of alluvial diamond mines known as wet diggings This designation distinguishes these secondary (placer) stream deposits from the dry diggings or pipe mines which were discovered shortly after 1868 In August 1870 the discovery of a 50-carat diamond in an intermittent stream led to the discovery of the first *kimberlite pipe This the Jagersfontein* Mine lies 80 miles east of Hopetown It is a nearly circular pipe with a cross sectional area of 25 acres and is the erosional remnant of the feeder neck of an old volcano The second pipe mine the Dutoitspan (60 acres) was discovered in the following month 20 miles southeast of Pneill where the *city of* Kimberley now stands The *Bultfontein* (62 acres) was discovered early in 1871 the DeBeers (43 acres) in May 1871 and the Kimberley (38 acres) the following month These last four and the city of Kimberley all lie within an area 3 miles in diameter The fifth member of the Kimberley group the Wesselton (49 acres) discovered in September 1890 originally called Premier lies 1¾ miles east of the Dutoitspan Of the many other pipes in this area only the Koffyfontein *between* Kimberley *and Jagersfontein has been* profitably operated

The only other profitably operated pipe mines are the Premier (80 acres) 70 miles northeast of Johannesburg discovered in 1903 the Williamson Mine (400 acres) at Mwadui Shinyanga District Tanganyika March 6 1940 the Bakwanga (80 acres) in the Belgian Congo 1949 and the Koidu pipe (½ acre) near Yengema Sierra Leone 1956

Little accurate information is available on the recently discovered kimberlite diamond pipes and associated stream deposits in the Yakutia district of Siberia They lie west of Yakutsk on the Lena River and north of Lake Baikal within the Arctic Circle They are within the permafrost region and the difficulties to be surmounted in successful mining are great It is doubtful if these diamond mines could be profitably operated in a free enterprise economy

Diamond recovery Ninety five per cent of the world s output of diamonds comes from Africa Most of the production is by large mining companies but a significant amount comes from individual operations from stream bed deposits in Sierra Leone Ghana and the Union of South Africa These small operations recover diamonds by a *method similar to the panning of gold from plac*ers taking advantage of the fact that the density of diamonds is greater than that of most other minerals Concentrates of the heavy minerals recovered by panning are hand picked for diamonds

Separation of concentrates The large mining companies which operate both alluvial (stream) deposits and pipe mines also use other methods for separating concentrates based on the higher density of diamonds The majority use large circular pans up to 16 ft in diameter for the first stage of recovery

A large horizontal wheel is supported above the pan by a shaft that passes up through an oversized cylinder in the center of the pan The top of this cylinder is lower than the outside walls of the pan Water and diamond bearing earth are continually fed into the pan and stirred by spokes extending downward nearly to the bottom of the pan from the rotating wheel The lighter material is carried by *the water over the rim of the central cylinder The* heavier minerals work slowly to the bottom outside edge of the pan where they are periodically removed The heavy minerals are further separated according to relative density by jigs which are known as pulsators in the diamond mining industry A more recent innovation is known
or float method in which a

density is made by mud of ferrosilicon flowing up ward in a large cone

Separation of the diamonds Diamonds are sep arated from the concentrates by hand sorting grease tables electrostatic methods fusion with alkalies surface tension or abrasion of the gangue minerals Grease tables are of two types (1) vibrat ing stationary tables coated with a layer of grease which is periodically removed and renewed by hand and (2) an endless belt to which a layer of grease is automatically applied before it moves across the table and is continuously removed after it leaves the table The concentrates fall onto the grease table and a stream of water carries the hydrophilic gangue across the table while the hy drophobic diamonds adhere to the grease

In electrostatic separation the concentrates are fed onto a grounded and rotating horizontal steel cylinder which lies beneath a strong electrostatic field Diamond is a nonconductor and retains the induced charge while the gangue minerals lose their charge to the grounded cylinders When the concentrates fall from the rotating cylinder they pass through a strong electric field of the same sign as the induced charge on the diamonds Small dia monds are deflected away from this second electric field to the far side of an adjustable knife edge which separates the falling diamond from the gangue

All the minerals in the concentrate have approxi mately the same density and all of these except diamond dissolve in molten alkali The chief draw back to this fusion process of separation is that the high temperature necessary may induce color changes in some diamonds

Because diamond is hydrophobic small crystals will float on water supported by surface tension An endless belt carrying the dry concentrates passes into a tank in which the water is slowly mov ing away from the point at which the belt enters The diamonds float while the hydrophilic gangue minerals sink

Before the treated concentrates are finally dis carded they may be ground in a ball mill The gangue minerals are reduced to a fine powder but any diamond that has not previously been removed resists the abrasion in the mill and can be recov ered by screening

Diamond cutting After recovery diamonds are referred to as rough Those of *gem quality* are called cuttable rough and all others are classed as industrial rough The poor grades of gem quality diamonds and finer industrials *are synonymous*

Inspection In cutting diamonds the objective is to obtain the max ...

... the stone if the sales price of a second qual ity stone of greater weight will be above the price of the smaller stone of first quality The rough crys-tal may be divided into two or more smaller stones whose total value will be greater than a single large stone Flaws may often be eliminated by this subdivision The most important step in diamond cutting is the decision as to how the stone will be cut

In subdividing an irregularly shaped crystal or eliminating flaws the stone may be either cleaved or sawed but only in certain crystallographic di rections There are four directions in which a dia mond may be cleaved parallel to any octahedron face and nine directions in which it may be sawed —three parallel to any cube face and six parallel to any dodecahedron face Some of the more expert cleavers can cleave parallel to the dodecahedron face but sawing in this direction is generally pre ferred

Cleaving In order to cleave a diamond a small groove is first cut in the edge of the diamond to be cleaved with the sharp edge of another diamond The cleaving iron (blade) is inserted in this groove parallel to the cleavage to be made, it is struck a sharp blow and the diamond breaks along smooth flat surfaces

Sawing Sawing is done with the edge of a rap idly rotating phosphor bronze disk that has been impregnated with diamond dust The starting saw which makes the first cut is thick enough to be rigid but after this initial groove a rapidly rotat ing paper thin phosphor bronze disk is used Initially the saw must be impregnated with dia mond dust but if properly started it continually re charges itself with material removed from the groove If the plane of the saw departs from paral lelism to the proper crystallographic direction (the sawing grain) progress is slower and ceases when the departure is 10–15° The saw must also rotate in the proper direction or no progress is made

Cutting If the finished diamond is to be round or oval in shape it goes to a man who is known in the ...

... les the

... ich like

... is held

against the rotating stone and "brutes" off the cor ner ...

... (holder) against a cast iron lap (skeif) that has been charged with oil and diamond dust The lap must move across the diamond in the proper crys tallographic direction or it will not cut The term cutting grain is used by diamond cutters to indicate the proper crystallographic direction for polishing and like the cleaving and sawing grains is some what analogous to the grain in wood Its direction may usually be determined from the external shape of the crystal or from markings on the crystal faces

Weight The weight of the finished gems is 50–60% of the weight of the rough stones if they are well formed crystals with few flaws The finished weight of irregularly shaped or badly flawed crys

tals is often very much less The metric carat 0 200 grams is the unit of weight by which both gem and industrial diamonds are sold A point is 1/100 of a carat and is used only in reference to gem diamonds

Industrial diamonds Industrial diamonds vary from the better grades which are identical with inferior gems to crushing bort which is suitable only for crushing to grit and powder sizes The better qualities are made into shaped diamond cutting tools or wire drawing dies Diamond cutting tools break on ferrous alloys because of the chattering which results from lack of ductility of the metals Shaped diamond tools have been supplanted in many of their former uses by sintered tungsten carbide

Tungsten carbide wire-drawing dies have also replaced diamond for the larger sizes of dies Diamond is still used for drawing wire smaller than 0 0025 in in diameter (average size of human hair) Tungsten filament wire for light bulbs and radio tubes is drawn at a low red heat and diamond is especially desirable because its hardness is little affected by these temperatures

For truing and shaping grinding wheels of alumina or silicon carbide diamond crystals in the shapes in which they are mined are used In most grinding and finishing operations it is only necessary to impart a smooth even finish to the wheel In form grinding the reverse of the shape to be ground is imparted to the grinding surface of the wheel Diamonds used for truing are usually called dressers and range in size from a fraction of a carat to several carats Those used for form grinding are referred to as thread grinders if in the original crystal form but if they have been shaped they may be called phonopoints because of their similarity to diamond phonograph needles Diamonds used for form grinding usually weigh less than 1/10 carat

Drill bort consists of crystals in their original form as mined which are mounted in the end of a cylinder called a bit or crown When the rotating cylinder is forced against a rock surface it wears its way into the rock That portion of the rock which extends through the cylinder up into the hollow drill rod which rotates the bit is called the core and is periodically removed A fluid usually water is circulated down the hollow drill rod and up the outside of the rod It cools and lubricates the drill bit and flushes out the rock particles abraded by the diamond The mining industry is the largest user of drill bort although this method of coring has been adapted to the testing of concrete and the foundations of buildings dams and bridges For these and mining purposes the average size of the diamonds used is less than 1/10 carat Large bits up to a foot in diameter are used by the oil industry both for coring and making hole These large bits use diamonds from 1/4 to 1 carat in weight Prior to World War II drill bort was the most important use of industrial diamonds

Beginning with World War II the greatly expanded use of tungsten carbide tools made bonded diamond grinding wheels by far the largest market for industrial diamonds A grinding wheel with a thin layer of imbedded diamond grit was developed to shape and sharpen these ultra hard tools Originally the manufacturer of these wheels crushed the bort and sized it but after World War II Industrial Distributors Ltd (1946) the sales outlet for industrial diamonds of the so called diamond syndicate processed the diamonds and sold the material as fragmented bort The annual world wide market before synthetic diamonds were developed totaled 12 000 000 carats three fourths of which was sold in the United States The mines at Bakwanga Belgian Congo produce nearly all of this the lowest grade of diamond

Synthetic or man made diamonds Many attempts had been made to manufacture diamonds prior to the announcement (Feb 15 1955) by the General Electric Company of their successful synthesis All claims of success prior to 1955 have been proven erroneous Synthetic diamonds are identical with natural diamonds in fundamental properties but differ in those characters that depend on the process of manufacture such as impurities size and shape Synthetic diamonds are made in the grit sizes (approximately 0 1 mm) These sizes are in greatest demand for the manufacture of bonded diamond grinding wheels for shaping and sharpening tungsten carbide tools This is the greatest single use of industrial diamonds and because of its importance in defense industries diamonds have been classed as a special strategic material Synthetics are superior to natural diamond for this use because they are single crystals roughly octahedral in shape with many cutting edges In crushing natural diamond many elongated slivers and flats are produced which reduce its efficiency

The small amount of impurity present in diamonds made by man does not reduce the hardness but does discolor many of the minute crystals It is possible to manufacture larger single diamond crystals a few millimeters in diameter but the cost is so great that they are not competitive with natural stones 0 5 mm in diameter or larger There seems little likelihood that single crystals of gem quality will be developed The higher pressures and temperatures in the range where synthetic diamonds are made give purer crystals Octahedrons also form under those conditions Dodecahedrons cubes and combinations of these form in lower pressure temperature ranges *See* CARBON [C B S]

Diapiric structures

Anticlines or domes in which a core of mobile rock has *broken through the overlying strata* In the process of concentric folding the rocks of the core occupy increasingly more limited space as the curvature of the fold becomes smaller If the rocks of the core are extremely plastic they may exert an

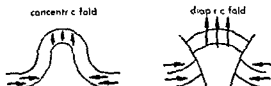


Fig 1 Horizontal and vertical movement in a concentric fold leading to diaprism (From L U de Sitter Structural Geology McGraw Hill 1956)



Fig 2 Penetration of slightly deformed strata by salt in the Hanoverian salt-dome region (After A Roll from L U de Sitter Structural Geology McGraw Hill 1956)

upward pressure sufficient to cause expulsion of the core through the crest (Fig 1) Diapiric structures may also form as a result of unequal loading by the sediments overlying a plastic layer The plastic rocks flow laterally away from places of high pressure and accumulate at places of low pressure Where accumulations of plastic rock reach a critical size the plastic material if it has a sufficiently low specific gravity compared to that of the surrounding rocks may penetrate the overlying rocks under the impetus of hydrostatic forces (Fig 2) *See* FOLD AND FOLD SYSTEMS SALT DOME [P H O]

Diarrhea

The passage of loose or watery stools usually at more frequent than normal intervals. Diarrhea is a symptom of many diseases and may be accompanied by nausea vomiting griping tenesmus and other general or specific indications of a disease

In simple acute diarrhea the causative agent is seldom known and the disorder spontaneously subsides in 2 or 3 days

The more common specific disorders which may produce diarrhea include intestinal infections such as dysentery cholera typhoid fever food poisonings and parasitic infestations food sensitivities drug and chemical irritation and vitamin deficiency states

Generalized toxic reactions produced by certain diseases such as measles thyrotoxicosis or pyogenic infections may also be accompanied by diarrhea as well as other symptoms

Emotional and psychic disturbances frequently produce diarrhea and other visceral derangements The poorly understood entities of regional ileitis and ulcerative colitis are perhaps related to these disturbances as are other psychosomatic disorders

Diarrhea is a common symptom in gastrointestinal obstruction or in inflammations from local infections or tumor invasion *See* BACILLARY DYSENTERY CHOLERA VIBRIO, FOOD POISONING BACTERIAL MEASLES PARASITOLOGY MEDICAL, TYPHOID FEVER VITAMIN [E G S T]

Diastem

A temporal break between adjacent geologic strata that represents nondeposition or local erosion but not a change in the general regimen of deposition (in contrast to unconformity) Diastems may be produced by the scouring action of shifting submarine currents which temporarily interrupt deposition on the continental shelf or by a shifting river within the deposits of its flood plain Or they may be produced simply by nondeposition in either environment of deposition where the absence of sediment reflects normal shifting of currents rather than an overall change in conditions *See* UNCONFORMITY

The existence of such breaks and their importance for interpreting the stratigraphic record was first pointed out by J Barrell (1917) Barrell showed that the time represented by the deposition of the beds actually observed in a stratigraphic sequence may be only a small fraction of the total time represented by the sequence as a whole even if the entire sequence was deposited under an essentially uniform regimen the rest of the time is represented by diastems It is now generally accepted that deposition on a shallow sea floor or on a flood plain is a discontinuous process *See* FLOOD PLAINS MARINE SEDIMENTS SEDIMENTATION (GEOLOGY) [J R]

Diastereoisomer

One of a pair of optical isomers which are not mirror images of each other (*see* OPTICAL ACTIVITY) A given diastereoisomer (or diastereomer) may not be optically active in which case it is an optically inactive meso form or it may be optically active in which case with its nonsuperimposable mirror image it constitutes an enantiomorphic pair of optical isomers Thus *m* tartaric acid is a diastereoisomer of both the *d* and *l* tartaric acids and it can form two optically active monoesters each of which is a diastereoisomer of monoesters of the *d* and *l* acids [W R Y]

Diastrophism

The general process or combination of processes by which the earth s crust is deformed also the results of this deforming action The term diastrophism was first used by J W Powell when in his study and discussion of major geologic features in the Cordilleran region of the United States he felt the need of a single word equivalent to the somewhat cumbersome phrase ' deformation of the earth s crust G K Gilbert a coworker adopted the new term as defined by Powell and suggested a dual subdivision of diastrophic processes and effects to distinguish the strong and comparatively localized deformation in mountain belts from the

simpler structural patterns of broad plateaus and basins that are bounded by zones of faulting and warping Gilbert approved the term orogeny (mountain making), already in general use for the more intensive deformation, and proposed epeirogeny (continent making) as the kindred term applying to simpler uplifts and depressions that affect wide segments of the crust

Diastrophism has operated continuously or repeatedly throughout geologic history Modern movements in disturbed zones have caused major earthquakes and measurable displacements of land surfaces Diastrophic effects in late geologic time are both topographic and structural In the Alps Andes, and other lofty mountains, layers of sedimentary rock that were formed on sea floors during the present geologic era are now at high altitudes and are much broken tilted and folded Similar structure in older formations now partly or completely leveled by erosion, marks locations of earlier mountain belts Dissection of these deformed belts has revealed not only buckling and fracturing but also large scale metamorphism of the sedimentary rocks The oldest deformed belts, widely exposed in areas of long continued erosion have bedrock consisting largely of crumpled and metamorphosed sedimentary strata partly engulfed in large bodies of granitic rock Thus intensive diastrophic action involves not only folding and fracturing but also metamorphism and major igneous activity The basic cause of diastrophism is unknown *See* Orogeny, Structural geology Tectonic patterns, Warping, earth crust [C R L]

Diatom

A markedly distinct group of algae belonging to the phylum Chrysophyta and the family Bacillariophyceae They have long been a favorite with naturalists because of the beauty of the symmetry and sculpturing of the siliceous cell walls They abound in all natural waters and are the basic food for many organisms The cells are unicellular although they may remain in loose chains or groups The cell wall formed in two halves which fit together like a shoe box and lid is largely composed of silicon dioxide and pectic compounds The markings on the valves are explained by the presence of minute thin places or pores in the walls In the different forms there is an almost endless diversity of patterns but the constant arrangement of mark

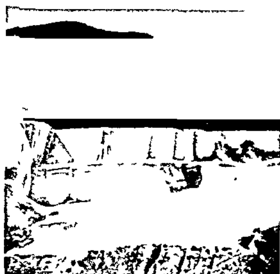
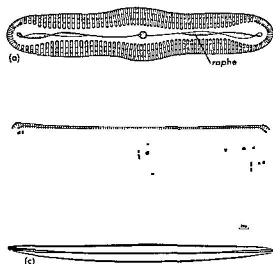


Fig 2 Diatomaceous earth deposit near Lompoc Calif (From H J Fuller and O Tippo College Botany rev ed, Holt, 1954)



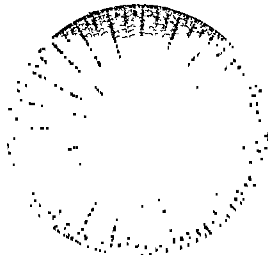


Fig 1 (a) Pinnularia, a diatom Top, or valve view, showing longitudinal slit or raphe (b) Side, or girdle view, showing overlapping halves or valves (c) Amphipleura pellucida, a bilateral diatom (d) Arachnoidiscus ehrenbergii, a concentric diatom (From H J Fuller and O Tippo, College Botany, rev ed, Holt, 1954)

ings in any given species forms the basis for delimiting the numerous species. The pennate diatoms exhibit a bilateral symmetry and are largely confined to fresh water (Fig. 1a, b, c). The centric diatoms have a radial symmetry and are usually marine (Fig. 1d). Reproduction is largely asexual, although sexual reproduction occurs in the formation of auxospores, which are zygotic in nature. Enormous beds of fossil diatoms, known as diatomaceous earth (Fig. 2), are found in various parts of the world (see DIATOMACEOUS EARTH). Because this material is inert chemically and has unusual physical properties, it is admirably suited for many scientific and industrial purposes, such as filtering agent, insulator against heat, cold, and sound, catalyst carrier, absorbent, filler, building material, abrasive, pharmaceutical preparations, and stratigraphic indicators. See CHRYSOPHYTA.

[P.A.]

Bibliography. See THALLOPHYTA.

Diatomaceous earth

Earth consisting of a friable, porous silica deposit made up of the opaline silica tests (shells) of diatoms. It is the dry, relatively unconsolidated equivalent of diatom ooze found on some parts of the sea floor today. It is usually white or cream colored. Diatomite is the indurated equivalent of diatomaceous earth; the pores are partially or completely filled with silica. See CHERT; DIATOM.

[R.S.]

Diatrymiformes

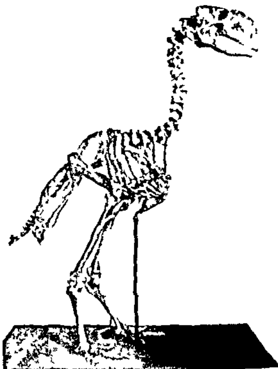
An order of extinct flightless birds known from early Eocene deposits of North America and Europe. Four genera are usually recognized in the family Diatrymidae and two in the family Gastornithidae, the latter known only from Europe. These were large, powerful birds with massive legs, relatively tiny wings, and large heads and beaks (see illustration). *Diatryma*, as much as 7 ft tall, was one of the l...

ming, is known from a nearly complete skeleton. This species, 2 meters tall, had strong legs, a massive body, and th...

...m high. The wings are tiny in comparison to the rest of the body, far too small to have had a volant function. The strong processes of the bones of head, neck, and legs indicate powerful muscular development and the probability of active predatory habits in securing food.

Among the species assigned to the family, three come from Wyoming, one from New Mexico, one from New Jersey, and one from Switzerland. The *Diatryma* group appears to have had remote ancestral connection with the flightless *Phororhacos* of southern South America, in which the larger species had a similarly heavy body form. Relationships of the order are uncertain, but it is placed near the cranelike birds. See AVES; AVES, FOSSILS; DINORNITHIFORMES; NEORNITHES.

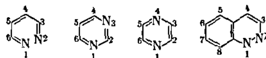
[A.W., K.C.P.]



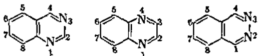
Reconstruction of *Diatryma steini*. Approximately 7 ft. (American Museum of Natural History)

Diazine

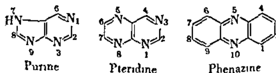
One of a group of organic heterocyclic compounds with a six-membered triunsaturated ring containing two nitrogen heteroatoms. See HETEROCYCLIC COMPOUNDS. The following formulations show the simple diazines, which include pyridazine, pyrimidine (see PYRIMIDINE), and pyrazine, as well as some bicyclic and tricyclic fused diazines.



Pyridazine or 1,2-diazine; Pyrimidine or 1,3-diazine; Pyrazine or 1,4-diazine; Cinnoline

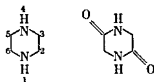


Quinazoline; Quinoxaline; Phthalazine



Purine; Pteridine; Phenazine

These compounds are basic and show some degree of aromatic character Several important natural products contain the pyrimidine purine pteridine and quinazoline ring systems Some synthetic dyes



Piperazine 2 5-Diketopiperazine

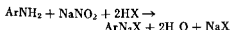
and medicinals are diazine derivatives Piperazine or hexahydropyrazine is a saturated cyclic dia mine basic enough to combine with carbon diox ide Cyclic self condensation of α amino acids gives 2 5 diketopiperazines [W J CE]

Bibliography A Albert The pteridines *Quart Revs* 6 197 237 1952 R C Elderfield *Hetero cyclic Compounds* vol 6 1957 J C E Simpson *Condensed Pyridazine and Pyrazine Rings* 1953 G A Swan and D G I Felton *Phenazines* 1957

Diazotization

The reaction between a primary aromatic amine and nitrous acid to give a diazo compound Diazo tization is important in organic chemical synthe sis First recognized by Peter Griess in 1858 the reaction is remarkable for the smoothness and completeness with which it can be carried out and for the reactivity of the products formed Its most striking use is in the large scale manufacture of the important azo class of dyes but it has likewise been invaluable in general synthesis in both chemi cal manufacturing and in research

Preparation of diazonium salts The most widely useful method of diazotizing a primary aromatic amine (the direct method) is by slow addition of an aqueous solution of sodium nitrite to a solution of the amine in dilute mineral acid held at 0–10°C Reaction is according to the following equation



$$ArNH_2 + NaNO_2 + 2HX \rightarrow ArN_2X + 2H\ O + NaX$$

Excess mineral acid usually about 2 5 moles is used to prevent formation of diazoamino by prod ucts and since the reaction is exothermic cooling is required to maintain a temperature at which the

sulfa
acid

solution the indirect method can be used in which a solution of sodium nitrite and the sodium salt of the aminosulfonic acid is run into a mineral acid solution containing ice For weakly basic amines such as 1 aminoanthraquinone diazotization is car ried out in concentrated sulfuric acid using nitro sylsulfuric acid as the diazotizing agent

Few primary aromatic amines resist diazotiza tion Not only can the amino groups on benzene naphthalene and their substituted derivatives be diazotized but heterocyclic amines such as amino thiazoles and aminopyridines will undergo the re action also Diamines in which the amino groups are on different aromatic nuclei in the same mole cule behave independently with respect to diazoti zation although the reaction can be carried out stepwise When the amino groups are on the same aromatic nucleus the reactivity of the second amino group to diazotization is lessened by the presence of the diazonium group first formed If the amino groups are ortho to one another a tria zole ring may result

When the diazotization reaction does fail it is usually because of oxidation or nitrosation caused by the nitrous acid Even when an aminophenol sensitive to oxidation is used diazotization has succeeded by special techniques For example 1 amino 2 hydroxynaphthalene 4 sulfonic acid can be diazotized in the presence of a copper salt

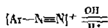
In most cases the diazo compound is used with out isolation If the solid diazonium salt is required for laboratory investigation diazotization can be carried out using an alkyl nitrite in alcohol or other inert solvent The diazonium salt precipitates as it forms or after addition of ether Commer cially diazonium salts useful for azoic color manu facture are isolated from aqueous diazotization media by precipitation as the free salt or as a less soluble complex salt for example with zinc chlo ride

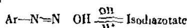
The free diazonium hydroxides are strong bases very soluble in water but insoluble in ether and other organic solvents They have limited stability and decompose in water at a rate accelerated by higher temperatures pH near neutral and the presence of such materials as finely divided metals In general electronegative substituents in the aryl nucleus increase the stability An outstanding characteristic of diazo compounds is their explosi bility particularly when dry Any isolated diazo compound should be handled with care even when damp There is however a tremendous difference in sensitivity among the various diazo compounds and for a given diazo salt anions such as chloride or sulfate result in a less dangerous compound than do anions such as nitrate chromate or perchlorate Diazo compounds are sensitive to light and this has led to their use in diazo type photocopying processes

Aliphatic diazo compounds are known although they are not important commercially Their struc ture and many of their reactions are different from those of the aromatic diazo compounds The sim plest diazomethane (CH_2N_2) finds use in the lab oratory as a methylating agent because it acts under very mild conditions Its reaction with acid chlorides is the basis of the Arndt Eistert method of increasing the carbon chain length of an ali phatic acid by one carbon Aliphatic diazo com pounds are usually made indirectly instead of by direct diazotization

The structure of aromatic diazo compounds has been the subject of extensive work Under acid

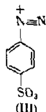
conditions they exist as diazonium salts (I) As the pH is raised this form is converted successively to the diazo hydroxide (II) and then to the isodiazotate

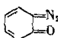


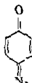


(I) (II)

Diazotized sulfanilic acid and similar compounds form inner salts (III) and diazotized o and p aminophenols form diazo oxides (IV) and (V) which are colored







(III) (IV) (V)

Coupling reactions Coupling of diazo compounds with phenols naphthols amines and certain other components to form azo dyes is the basis of their most important commercial use Annual

{

of diazo compounds to give azo dyes is by attack of the strongly electrophilic diazonium ion (I) on a position of high electron density in a nucleophilic coupling component With phenols the greater electron density at the para position in the phenolate ion aids coupling unless the pH becomes so high that the diazonium salt is substantially converted to the noncoupling isodiazotate Coupling with amines occurs more readily near the neutral point than at a lower pH where amine salt is present

Electronegative substituents in the diazo compound increase its coupling power and the diazo compound of trinitroaniline which is the most reactive diazo compound in this respect will couple even with mesitylene Besides coupling with amines and phenols diazo compounds can couple with phenol and naphthol ethers compounds containing reactive methylene groups such as acetoacetanilides or nitroparaffins and certain unsaturated hydrocarbons

Although phenols couple almost exclusively at an open para pos t ...

c

t .

ing the coupling positions may be displaced By use of an excess of diazo compound two or even three diazo molecules can enter a coupling component such as phenol The second and third couplings occur with difficulty however and the amount of the third coupling is not great even under the most favorable conditions. Coupling with phenol or naphthol ethers occurs more slowly than when the free hydroxyl group is present and the reaction is frequently accompanied by at least partial dealkylation

Amines couple more slowly than phenols and with aniline only one diazo substitution occurs By control of conditions a molecule containing both hydroxyl and amino groups can be caused to couple selectively at a position activated by one but not the other of these groups For example H Acid (8 amino-1 naphthol 3 6 disulfonic acid) couples ortho to the hydroxyl group under slightly alkaline conditions and ortho to the amino under slightly acid conditions and two different diazo compounds may react with the molecule In such cases the acid coupling is always made first

Various solvents of which aqueous pyridine is the most common are often used to increase the rate of slow couplings

Reduction and displacement reactions Diazo compounds are reduced to substituted hydrazines when treated with sodium sulfite Aryl hydrazines thus made available were of inestimable value in structure determination in sugar chemistry and today serve as intermediates for the manufacture of pyrazolones important in dye chemistry Milder reduction using ethyl alcohol sodium stannite or hypophosphorous acid replaces the nitrogens with hydrogen (deamination) Oxidation is rarely used but can be carried out by use of reagents such as sodium hypochlorite to give nitroamino compounds

Of great synthetic value is the replacement of the diazonium group by halogens or the cyano group This reaction is carried out by treating the diazo compound with for example hydrochloric acid in the presence of cuprous chloride (Sandmeyer reaction) The Gatterman modification of the Sandmeyer reaction uses finely divided copper instead of the cuprous salt Heating the diazo compound in concentrated hydrofluoric acid will replace the nitrogens with fluorine and this is a technical method for making fluorinated aromatic compounds Heating the isolated diazonium fluoborate also introduces fluorine into the molecule (the Schiemann reaction) See HALOGENATED HYDROCARBON

Replacement of the diazo group with sulfur by treatment with sodium polysulfide or sodium xanthate is of technical importance in the synthesis of thioindigoid dyes Heating the diazo compound with dilute acid replaces the nitrogens by hydroxyl and heating with alcohols can give substitution by alkoxyl groups The latter is often a side reaction in deamination by means of ethyl alcohol Aryl arsonic acids are formed by treating the diazo compound with trivalent arsenic compounds such as sodium arsenite (Bart reaction) Similar antimony containing compounds can be made The double salts of diazonium halides with those of mercury tin and lead can be reduced to give organometallic compounds having carbon attached directly to metal in the position previously occupied by the diazo group

Symmetrical biaryls are formed by treatment of diazo compounds with a reducing agent such as the ammoniacal cuprous ion This reaction is of technical importance in the preparation of (1 1 binaphthalene) 8 8 dicarboxylic acid a vat dye intermediate Unsymmetrical biaryls can be made by the Gomberg Bachmann reaction which is carried out by causing the aryl diazo hydroxide to react with an aromatic liquid Intramolecular arylation by elimination of the diazo group is called the Pschorr reaction For example diazotized 2 amino α phenylcinnamic acid when treated with copper powder yields phenanthrene 9 carboxylic acid See ANILINE DYE [W A F]

Bibliography R A Adams (ed) *Organic Reactions* vol 2 1944 P H Groggins (ed) *Unit Processes in Organic Synthesis* 5th ed 1958 K H Saunders *The Aromatic Diazo Compounds* 2d ed 1949 K Venkataraman *The Chemistry of Synthetic Dyes* vol 1 1952

Dibranchia

A subclass of the Cephalopoda which contains the extinct Belemnoidea and all living cephalopods with the exception of *Nautilus* All members of this group possess two gills and when present the shell is internal They are presumed to have arisen from the Nautiloidea and the first fossil dibranchiate *Eobelemnites* is from the Mississippian The belemnites have a straight internal shell which is hollow and bears septa but otherwise closely resembles the structure of the internal shell of *Spirula* and *Sepia* The belemnites arose in the Upper Mississippian and disappeared at the close of the Cretaceous *Belemnites* is an index fossil of Jurassic and Cretaceous rocks throughout the world

Decapoda The decapods arose in the Jurassic A sepioid *Voltzia palmeri* occurs in the Upper Jurassic of Cuba and a teuthoid *Plesioteuthis* also is known from this period Two basic divisions occur the sepioids (containing the bathypelagic genus *Spirula* with a coiled internal shell) and a large number of small squat broad finned sepiolids which are chiefly bottom dwellers in deeper waters and the larger free swimming squids The last present some of the most bizarre forms to be found in the class See DECAPODA (MOLLUSCA)

Octopoda The octopods have fleshy suckers but in the squids the suckers are equipped with chitinous rings which may be smooth or toothed along the margin and in some they are modified to form long sharp claws for capturing prey In some the hooks have a swivel like attachment and the claws may be sheathed when not in use

In the bathypelagic species complex light organs are found While their use has not been definitely established they may be for recognition or aid in capturing prey They occur on the mantle or internally on the viscera and on the head arms eyes and tentacles Several colors may be emitted The source is luciferin luciferase usually with the organs composed of light source reflector diaphragm lens and color filter One spectacular species is *Oregoniateuthis springer* Voss from the Gulf of Mexico which contains over 34 large photophores See BIOLUMINESENCE OCTOPODA

Vampyromorpha In 1940 S Pickford proposed that the Vampyromorpha be raised to equivalent rank with the Decapoda and the Octopoda It is represented by *Vampyroteuthis infernalis* dwelling in the deeper waters of tropical and temperate seas It bears eight arms and two vestigial retractile filaments an uncalcified gladius and as an adult has one pair of paddle shaped fins It is considered by some to be an intermediate form between the decapods and the octopods There is no fossil record

Protective mechanisms Most dibranchiates emit ink from a special gland Contrary to public opinion in most species the ink discharged is not a smoke screen but coagulates in the water as an object of the approximate size of the animal This is seized by the enemy while the colorless cephalopod darts away In some deep sea forms the ink sac is lost while in others luminous ink is discharged [C L V]

Bibliography Frank W Lane *Kingdom of the Octopus the Life History of the Cephalopoda* 1957

Dichroism

Certain anisotropic materials have different absorption coefficients for light polarized in different directions Such materials are termed dichroic and the property is called dichroism

There are few natural materials which exhibit strong dichroism One of the first to be discovered was tourmaline Light transmitted by thin plates of dark forms of tourmaline is almost completely polarized See POLARIZED LIGHT

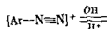
In isotropic optical materials the optical density is defined as

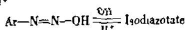
$$d = \log \frac{I_0}{I}$$

where I_0 is the intensity of the incident light and I that of the transmitted light In anisotropic materials that are dichroic the value of d can vary as a function of the vibration direction of the electric vector of the light wave Just as the index ellipsoid is used to define the birefringence of a material a density surface can be used to define the dichroism See CRYSTAL OPTICS

Compared to the literature on birefringence and optical activity there has been relatively little material on dichroism This is partly because of the difficulty in making measurements The Kramers Kronig relationship shows that any material whose refractive index is different from unity and varies as a function of wavelength will absorb radiation at some wavelength From the Kramers Kronig relationship it is apparent that all optically anisotropic materials should be dichroic From the values of the refractive index at different wavelengths the spectral positions and intensity of the absorption can be calculated In a linear birefringent

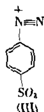
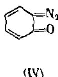
conditions they exist as diazonium salts (I) As the pH is raised this form is converted successively to the diazo hydroxide (II) and then to the isodiazotate





(I) (II)

Diazotized sulfanilic acid and similar compounds form inner salts (III) and diazotized *o* and *p* aminophenols form diazo oxides (IV) and (V) which *are colored*


(III) (IV) (V)

Coupling reactions Coupling of diazo compounds with phenols naphthols amines and certain other components to form azo dyes is the basis of their most important commercial use Annual production of this type of dye in the United States is about 55 000 000 lb with a value of $75 000 000

The most probable mechanism for the coupling of diazo compounds to give azo dyes is by attack of the strongly electrophilic diazonium ion (I) on a position of high electron density in a nucleophilic coupling component With phenols the greater electron density at the para position in the phenolate ion aids coupling unless the pH becomes so high that the diazonium salt is substantially converted to the noncoupling isodiazotate Coupling with amines occurs more readily near the neutral point than at a lower pH where amine salt is present

Electronegative substituents in the diazo compound increase its coupling power and the diazo compound of trinitroaniline which is the most reactive diazo compound in this respect will couple even with mesitylene Besides coupling with amines and phenols diazo compounds can couple with phenol and naphthol ethers compounds containing reactive methylene groups such as acetoacetanilides or nitroparaffins and certain unsaturated hydrocarbons

Although phenols couple almost exclusively at an open para position coupling will occur in the ortho position if the para position is blocked In the more reactive diazo compounds groups blocking the coupling positions may be displaced By use of an excess of diazo compound two or even three diazo molecules can enter a coupling component such as phenol The second and third couplings occur with difficulty however

when the free hydroxyl group is present and the reaction is frequently accompanied by at least partial dealkylation

Amines couple more slowly than phenols and with aniline only one diazo substitution occurs By control of conditions a molecule containing both hydroxyl and amino groups can be caused to couple selectively at a position activated by one but not the other of these groups For example H Acid (8 amino-1 naphthol 3 6 disulfonic acid) couples ortho to the hydroxyl group under slightly alkaline conditions and ortho to the amino under slightly acid conditions and two different diazo compounds may react with the molecule In such cases the acid coupling is always made first

Various solvents of which aqueous pyridine is the most common are often used to increase the rate of slow couplings

Reduction and displacement reactions Diazo compounds are reduced to substituted hydrazines when treated with sodium sulfite Aryl hydrazines thus made available were of inestimable value in structure determination in sugar chemistry and today serve as intermediates for the manufacture of pyrazolones important in dye chemistry Milder reduction using ethyl alcohol sodium stannite or hypophosphorous acid replaces the nitrogens with hydrogen (deamination) Oxidation is rarely used but can be carried out by use of reagents such as sodium hypochlorite to give nitroamino compounds

Of great synthetic value is the replacement of the diazonium group by halogens or the cyano group This reaction is carried out by treating the diazo compound with for example hydrochloric acid in the presence of cuprous chloride (Sandmeyer reaction) The Gatterman modification of the Sandmeyer reaction uses finely divided copper instead of the cuprous salt Heating the diazo compound in concentrated hydrofluoric acid will replace the nitrogens with fluorine and this is a technical method for making fluorinated aromatic compounds Heating the isolated diazonium fluoborate also introduces fluorine into the molecule (the Schiemann reaction) *See* HALOGENATED HYDROCARBON

Replacement of the diazo group with sulfur by treatment with sodium polysulfide or sodium xanthate is of technical importance in the synthesis of thioindigoid dyes Heating the diazo compound with dilute acid replaces the nitrogens by hydroxyl and heating with alcohols can give substitution by alkoxyl groups The latter is often a side reaction in deamination by means of ethyl alcohol Aryl arsonic acids are formed by treating the diazo compound with trivalent arsenic compounds such as sodium arsenite (Bart reaction) Similar antimony containing compounds can be made The double salts of diazonium halides with those of mercury tin and lead can be reduced to give organometallic compounds having carbon attached directly to metal in the position previously occupied by the diazo group

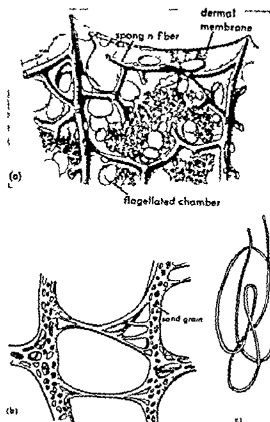


Fig 1 Skeletal features of Dictyocerat da (a) Sect on through a bath sponge Spongia off cinal s (after Schulze 1879) (b) Port on of skeleton of Irc n a (c) Flament of Irc n a (after Lendenfeld 1889)

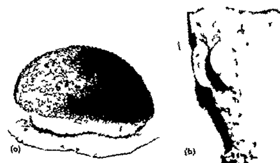


Fig 2 (a) Spong a off cinalis (after Schulze 1879) (b) Phyllospong a papyracea (after Lendenfeld 1889)

Spongudae small spherical flagellated chambers are present and these characteristically join the ex halant canals by way of narrow channels In the family Dysideidae there are larger sac shaped chambers which join the exhalant chambers di rectly A leathery dermis probably reinforced with spongin is typically present and is often beset with cone shaped elevations marking places at which fibers reach the surface The genus *Ircinia* is char acterized by the presence throughout the flesh of numerous thin filaments with terminal kno The nature and function of these filaments are unknown

Dictyoceratida are mostly of considerable size and form massive lobate or branching colonies Some are leaflike in shape or vase-shaped *Crypto spongia enigmatica* is discoidal in shape and has a long slender stalk

Dictyoceratid sponges are most abundant in tidal and shallow waters of tropical and subtropical re gions Relatively few species occur in Arctic and Antarctic seas and most occur above the continen tal slope A few species are found down to 1000 meters and at least one *Cryptospongia enigmatica* occurs at a depth of 2000 meters in the Indian Ocean *See* DEMOSPONGIAE [W D H]

Dicyemida

An order of Mesozoa comprising minute worm like parasites of the renal organs of cephalopod mollusks They are composed of a single layer of large ciliated epithelial cells enclosing one or more elongate axial cells Each axial cell contains, in addition to its own nucleus, reproductive cells and developing larvae These larvae when fully de veloped escape from the parent organism

There are two reproductive phases in which the reproductive individuals are termed the nemato gens and rhombogens respectively Ordinarily they contain only one axial cell In some species the first nematogens found in very young cephalo pods (presumably the infecting agent) have two or three axial cells in linear series as well as other differences These have been termed stem nemato gens They give rise to and are soon replaced by the ordinary or primary nematogens as the popu lation of parasites increases

During the nematogen phase vermiform larvae are formed asexually from the germ cells, or axo blasts in the axial cell They increase the infec tion in the same cephalopod This is succeeded by the rhombogen phase in which infusoriform or swarm larvae are liberated These are discharged into the sea with their host s urine and their fate is unknown

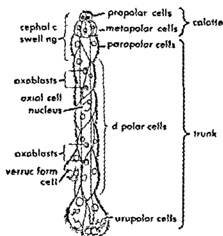


Fig 1 General zed drawing of a young d cy

In the rhombogen phase the axoblasts do not form the infusoriform directly. Most degenerate but a few give rise to infusorigens. The infusorigen consists of an axial cell containing sperm cells in various stages of development surrounded by peripheral cells which develop into egg cells. As these ripen and are fertilized, they detach from the infusorigen and lie free in the axial cell of the parent rhombogen. Here they undergo rapid cleavages forming infusoriform larvae.

Zoologists are not in agreement regarding many points in the life cycle, particularly in respect to the interpretation of the infusorigen and its products.

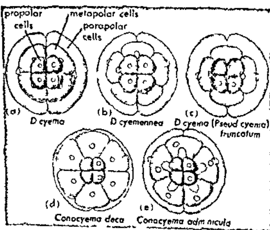


Fig 2 Diagrammatic representation of various arrangements of the cells of the calotte in dicyemids (a) Genus *Dicyema* (b) Genus *Dicyemennea* (c) *Dicyema (Pseudicyema) truncatum* (d) *Conocyema deca* (e) *Conocyema adminicula* young nematogen showing anterior end of axial cell just beginning to branch into the large metapolar cells

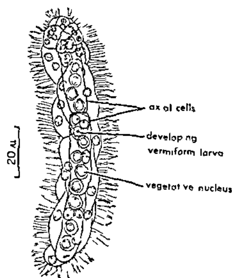


Fig 3 Young stem nematogen of *Dicyema schulzianum* (After H Nouvel)

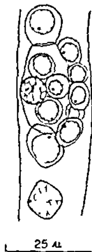


Fig 4 Young infusorigen in axial cell of a rhombogen optical section

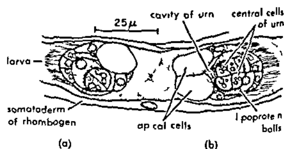


Fig 5 Infusoriform larvae in axial cell of a rhombogen (a) parasagittal and (b) frontal views (After B H McConnaughey, 1941)

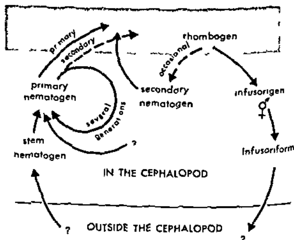


Fig 6 Diagram of the life cycle of the dicyemids A single arrow represents a new generation arising from germ cells within the preceding A dashed arrow indicates merely the transformation of an individual from one phase to another Parenthesis indicates that the form enclosed does not leave the axial cell of the individual in which it arose

A few small aberrant species, the heterocyemids possess atypical calottes and often a reduced trunk In some the axial cell may branch and the somatic cells may form a syncytium in which cell boundaries are obscure or absent The vermiform larvae differ in appearance from those of typical dicyemids That of *Conocyema* (*Microcyema*) *vespa* is known as Wagener's larva The so called heterocyemids are grouped into a subfamily Conocyeminae of the family Dicyemidae *See* ACTINOMYXIDIA CEPHALOPODA [B H M]

Bibliography B H McConnaughey The life cycle of the dicyemid mesozoa *Univ Calif (Berkeley) Publs Zool*, 55(4) 295–336 1951 H Nouvel Les dicyemides *Arch biol Liege* 58(1–2) 59–220 1947 and 59(2) 147 221 1948

Dielectric constant

For a given dielectric material the ratio of the capacitance of a capacitor which has the region between its plates filled with the given dielectric to the capacitance of the same capacitor when the region between its plates is a vacuum is known as the dielectric constant The defining equation is

$$\kappa = C/C_0$$

where C is the capacitance of the dielectric filled capacitor and C_0 is the capacitance of the empty capacitor The dielectric constant κ is also known as the specific inductive capacity or the relative permittivity It is perhaps most familiar as the proportionality constant in Coulomb s law of electrostatics For a given charge distribution the dielectric constant expresses the ratio of electric field strength in vacuum to that in a dielectric the latter field being reduced by the polarization of the dielectric medium *See* CAPACITANCE CAPACITOR COULOMB S LAW DIELECTRICS, ELECTRIC FIELD PERMITTIVITY

The values of κ for the frequency or static fields range from 1 to over 10 000 for typical dielectrics The dielectric constants of gases are only slightly greater than unity while high values occur for many polar liquids and certain ionic solids The table lists some representative values

Measurement The experimental methods of measuring dielectric constants depend on the frequency range under investigation For frequencies below about 10^9 cps the permittivity or impedance of a dielectric sample inserted in a parallel plate capacitor may be measured in suitable circuits a Schering bridge arrangement is commonly employed up to 10^7 cps and resonant circuits in the range 10^4–10^9 cps For frequencies above 10^9 cps the dielectric constant may be determined by measuring the interaction of electromagnetic waves with the medium From about 10^8 to 10^{11} cps the material is usually inserted in wave guides or coaxial lines and the standing wave patterns measured At still higher frequencies optical techniques involving reflection and transmission measurements are employed Measurement techniques employed in analytical chemistry are discussed later

Selected dielectric constants

Substance	Temperature °C	Frequency cps	κ
Dry air CO_2 free	20		1 00054
NaCl	25	10^2	5 9
MgO	25	10^2	9 65
Al_2O_3	25	10^2	10 55[a]
			8 6[b]
TiO_2	25	10^5	170[a]
			86[b]
$BaTiO_3$	25	10^6	180[a]
			2000[b]
Polyethylene and paraffin	25	10^2	2 25
Rubber vulcanized	25	10^2	2 94
Quartz fused	25	10^2	3 78
Mica	25	10^4	7 3
			6 9[d]
Water	25	10^5	78 5
Ice	12	10^6	4 8
HCN	20		114 9
CH_3OH	25	10^6	31

[a] Electric field parallel to principal axis of crystal
[b] Electric field perpendicular to principal axis of crystal
[c] Electric field parallel to sheet of material
[d] Electric field perpendicular to sheet of material

Macroscopic theory The dielectric constant κ is a dimensionless parameter relating the macroscopic quantities displacement $\mathbf{D}$ and polarization $\mathbf{P}$ with electric field $\mathbf{E}$

$$\kappa = \frac{\epsilon}{\epsilon_0} = \frac{\mathbf{D}}{\epsilon_0 \mathbf{E}} = 1 + \frac{\gamma \mathbf{P}}{\epsilon_0 \mathbf{E}} = 1 + \gamma\chi$$

where ϵ and ϵ_0 are the permittivities of dielectric and vacuum respectively χ is the electric susceptibility and γ is a geometrical factor ($\epsilon_0 = 1$ and $\gamma = 4\pi$ in cgs electrostatic units $\epsilon_0 = 8\,854 \times 10^{-12}$ farad/m and $\gamma = 1$ in rationalized mks units) The quantities κ and χ are scalar or tensor quantities for isotropic or anisotropic dielectrics respectively *See* SUSCEPTIBILITY ELECTRIC

For electric fields varying sinusoidally with time where the phase of the displacement may be retarded with respect to the field the preceding equation may be employed using complex number notation Thus $\kappa^* = \kappa' - i\kappa''$ where κ' and κ'' designate the components of the permittivity ratio in phase with E and retarded ¼ cycle respectively The term dielectric constant is usually restricted to the real part κ' while κ^* is designated as the complex relative permittivity

Microscopic theory The dielectric constant of a material depends on its polarization in an applied field or microscopically on the relative displacements in the field direction of the electrons and nuclei comprising the molecules of the dielectric These displacements are associated with changes in rotational and vibrational motions of the electrons and nuclei upon application of an electric field This leads to a frequency and temperature dependence resulting from the inertial characteris

tics of the motions and the initial state of excita tion of the system The dielectric constant also de pends on field strength since for sufficiently high fields the polarization will no longer be propor tional to the field because saturation or breakdown phenomena occur

Theory of static polari ation A molecule in an electric field develops an average dipole moment $(\mu)_{av} = \alpha E$ where α is the polarizability For a polar molecule α may be expressed as $\alpha = \alpha_o + \alpha_p$ where α_p is the polarizability arising from the partial orienta tion of the permanent dipole moment μ_p and α_o is the polarizability due to all other processes Then $(\mu)_{av} = \alpha_o E + (\mu_p \cos\theta)_{av}$ where $(\mu_p \cos\theta)_{av}$ is the average value of the component of μ_p in the field direction θ being the angle between μ_p and F

The potential energy U of the permanent dipole is given by $U = -|\mu_p||E| \cos\theta$ For a system in thermodynamic equilibrium with the probabilities of the possible states given by the Boltzmann distribu tion it can be shown that $(\mu_p \cos\theta)_{av} = \mu_p L(x)$ where the Langevin function $L(x) = \coth x \quad (1/x)$ and $x = |\mu_p||F| \; kT$ (where k is the Boltzmann con stant and T is the absolute temperature) For $x \ll 1$ $L(x) \cong x/3$ in this approximation where the micro scopic parameters have been multiplied by N the number of molecules per unit volume to give the polarization

$$P = N(\mu)_{av} = N(\alpha_o + |\mu_p|^2/3kT)F$$

This is a form of the Langevin Debye formula from which the permanent dipole moments of polar mol ecules may be obtained from the temperature varia tion of measured values of polarization For addi tional information *see* MOLECULAR STRUCTURE AND SPECTRA

Local field In the previous section the mutual interaction of the molecules was not considered ex

local field E_l the average electric field strength at

charges and polarization by molecules outside the sphere the sphere is taken sufficiently large that the medium outside may be considered continuous E_r is obtained by subtract ... the field F

l and the normal to the surface The field at the center contributed by a unit area of the surface is equal to $-\gamma s/4\pi\epsilon_0 R$ where R is the radius of the sphere and its component in the direction of P is $-\gamma s \cos\theta/4\pi\epsilon_0 R^2$ Integrating over the surface gives the net field $E_p = \gamma P/3\epsilon_0$ Finally

$$E_r = E - E_p = E + \frac{\gamma P}{3\epsilon_0} = \frac{(D - 2\gamma P/3)}{\epsilon_0}$$
$$= \frac{E(\kappa' + 2)}{3} = F(1 + \gamma\chi/3)$$

By assuming that $E_s = 0$, one obtains the Lorentz local field $E_l = E_r$ which leads to the following ex pressions for P, χ and κ'

$$P = N(\mu)_{av} = N\alpha E_l = N\alpha(E + \gamma P/3\epsilon_0)$$
$$= N\alpha E/(1 - \gamma N\alpha/3\epsilon_0)$$
$$\chi = P/\epsilon_0 E = N\alpha/\epsilon_0(1 - \gamma N\alpha/3\epsilon_0)$$
$$\kappa' = 1 + \gamma\chi = 1 + \gamma N\alpha/(\epsilon_0 - \gamma N\alpha/3)$$

The last equation may be solved for α giving $\gamma N\alpha/3\epsilon_0 = (\kappa' - 1)/(\kappa' + 2)$ which is a form of the Clausius Mosotti equation This formula gives fairly good agreement with experiment up to moderate densities for nonpolar molecules but fails to account for the behavior of strongly polar molecules

In the Lorentz field approximation, the suscepti bility becomes infinite if $\gamma N\alpha = 3\epsilon_0$ Since α for polar molecules is given by $\alpha = \alpha_o + |\mu|^2/3kT$ there exists a critical temperature below which $\alpha > 3\epsilon_0/\gamma N$ and the material should undergo spontaneous polariza tion This so called "polarization catastrophe is not confirmed by experiment except for a class of crystals known as ferroelectrics (*see* FERROELECTRICS)

Onsager theory The polarization catastrophe is avoided in a dielectric theory of polar molecules due to L Onsager In his treatment the local field is calculated for an actual spherical cavity of molecular size in the dielectric using Laplace's equation which gives $E_l = 3\kappa E/(2\kappa' + 1)$ A polar molecule with dipole moment μ_p inserted in the cavity induces an additional reaction field

$$E_R = \frac{\gamma\mu_p 2(\kappa' - 1)}{\epsilon_0 V 3(2\kappa' + 1)}$$

where V is the volume of the cavity This reaction field is parallel to μ_p and thus exerts no aligning torque on the molecule For this theory, the polariza tion is given by

$$P = N\alpha E_l = N\alpha \frac{3\kappa' E}{2\kappa' + 1}$$

Since $\gamma P = (\kappa' - 1)\epsilon_0 E$ one has

$$\frac{\gamma N\alpha}{\epsilon_0} = \frac{(2\kappa' + 1)(\kappa' - 1)}{3\kappa'}$$

and $$\kappa' = \frac{1}{4}\left[1 + 3z + 3\left(1 + \frac{2z}{3} + z^2\right)^{1/2}\right]$$

where $z = \gamma N\alpha/\epsilon_0$ The agreement with theory is satisfactory for most polar liquids but is inade quate for systems with hydrogen bonds [R D W]

Applications in analysis The dielectric constant because it is related to chemical structure can be used for both qualitative and quantitative analysis Alone it is a nonspecific indication of qualita tive constitution Unless the sample is known to be

a pure material the dielectric constant is of very little value since for mixtures it is generally not a simple additive function of composition Even for pure samples it is seldom used because there are generally more sensitive methods available On the other hand from measurements of the dielectric constant and refractive index taken together it is possible to compute the dipole moment of a species This quantity is a measure of the separation of charge within the molecule and thus often gives explicit clues concerning structure when composition is known from other means Azobenzene ($C_{12}H_{10}N_2$) diiodoacetylene (C_2I_2) and carbon suboxide (C_3O_2) were first assigned their symmetrical molecular structures unambiguously because dielectric measurements indicated a lack of dipole moment The α and β isomers of benzene hexachloride ($C_6H_6Cl_6$) used in insecticides were first characterized in the same fashion

The dielectric constant is a nonadditive function for mixtures and so for quantitative analysis it is necessary to prepare calibration curves relating measurements to composition for standard samples After this has been done however the method is rapid and efficient It is most applicable to two component mixtures When the ratio of the dielectric constants of the components of a two component mixture is 2 1 the accuracy of the determination is 0 2 2% The method can be applied to three component mixtures also if some independent method is available for determining one of the components For more complex systems the errors mount rapidly except in special cases If the dielectric constants for all but one constituent in a multicomponent system are similar and there is little interaction between them in solution then the unique component can often be determined This is the situation in the analysis for toluene in the presence of complex mixtures of aliphatic hydrocarbons in petroleum refining Determination of moisture in cereal grains and other solids is based on a similar principle

Experimentally the dielectric constant is determined by the ratio of the capacitances of a capacitor measured with and without the sample between its plates Measurements can be made at low (1000–10 000 cps) or high (1 Mc) frequencies The equipment needed at low frequencies is simpler to operate than that needed at high frequencies However direct contact between sample and electrodes is unnecessary at high frequencies See DIPOLE MOMENT POLARIZATION (DIELECTRICS) [W H R]

Bibliography C P Smyth *Dielectric Behavior and Structure* 1955 A R von Hippel *Dielectrics and Waves* 1954

Dielectric heating

The heating of a nominally electrical insulating material due to its own electrical (dielectric) losses when the material is placed in a varying electrostatic field

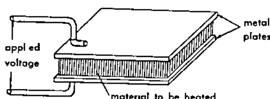


Fig 1 Basic assembly for dielectric heating

The material to be heated is placed between two metal plates called electrodes as in Fig 1 The electrodes are connected to a source of 2–90 megacycles produced by a high frequency oscillator

The resultant heat is generated within the material and in homogeneous materials is uniform throughout It is a rapid method of heating and is not limited by the relatively slow rate of heat diffusion present in conventional heating by external surface contact or by radiant heating For a discussion of dielectric losses see DIELECTRICS

Applications This technique is widely employed industrially for preheating in the molding of plastics for quick heating of thermosetting glues in cabinet and furniture making for accelerated jelling and drying of foam rubber foundry core baking and drying of wall board and other products Its advantages over conventional methods are the speed and uniformity of heating which offset the higher equipment costs Because of the absence of high thermal gradients an improved end product quality is usually obtained Figure 2 shows a 30 megacycle unit for preheating plastics prior to molding

Dielectric heating process The heating rate obtainable in watts is

$$P = \frac{1\,414 A f E^2 \epsilon \tan \delta \times 10^{-12}}{d}$$

where P = heating rate watts
A = material area in ^{2}
d = thickness of material or electrode spacing in
f = frequency cps
E = voltage across the electrodes rms
ϵ = dielectric constant of material
δ = power factor

The factor $\epsilon \tan \delta$ is referred to as the loss factor and serves to indicate the relative rates of heating for different materials at the frequency to be employed in high frequency fields of the same intensity For values and discussion of ϵ see DIELECTRIC CONSTANT

The controllable variables in the process are voltage E and frequency f Voltage can be raised only within limits determined by corona or dielectric breakdown generally not exceeding 15 000–20 000 volts Voltage gradient across the work is generally the principal determinant and this ranges between 1500 and 5000 volts per inch depending upon the porosity of the work See CORONA DISCHARGE

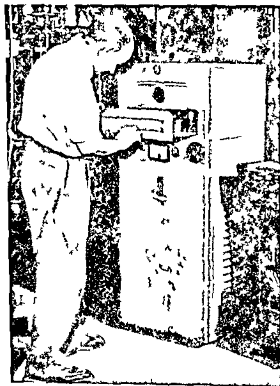


Fig 2 Dielectric heating unit for preheating plastics The material to be heated is loaded on the sliding drawer in front When closed this drawer places the material between the work electrodes (Chemetron Corp)

Frequencies used are as high as practicable so that voltages can be kept within the limits given Most standard equipment uses frequencies from 5 to 40 megacycles with power outputs up to 125 kilowatts Power rating decreases for the higher frequencies

The work electrodes and the material being heated function like a capacitor They are usually connected as a part of a resonant circuit which is tuned to the oscillator frequency to obtain maximum transfer of power The heating elements are enclosed in shielded cages or within conveyorized [illegible]

radio communication channels industrial radio frequency (rf) heating installations operating above 10 000 cycles must be properly shielded

The usual method is to enclose completely all rf circuits and electrodes in a metal cabinet using wire mesh shielding over ventilation openings and good electrical contacting or overlapping surfaces on access doors Where inlet and exit openings must exist as in conveyorized installations a metallic shield in the form of a vestibule or throat is used The length and area of the shield are designed to attenuate rf radiation from the electrode area

[C P B]

Bibliography G H Brown C N Hoyler and R A Bierwirth *Theory and Application of Radio Frequency Heating* 1947 J W Cable *Induction and Dielectric Heating* 1954 Recommended Practices for Minimization of Interference from Radio Frequency Heating Equipment AIEE Standard 951 New York

Dielectrics

Materials which are electrical insulators or in which an electric field can be sustained with a minimum dissipation of power In a more general sense dielectrics include all materials except condensed states of metals *See* INSULATOR ELECTRIC

Dielectrics are employed as insulation for wires cables and electrical equipment as polarizable media for capacitors in devices used for the propagation or reflection of electromagnetic waves and for a variety of dielectric devices such as rectifiers and semiconductor devices piezoelectric transducers dielectric amplifiers and memory elements

The electrical response of a dielectric can be described by its dielectric or breakdown strength conductivity or dielectric loss and dielectric constant

Dielectric strength This is defined as the maximum electric field which can be applied to a dielectric without causing breakdown the abrupt irreversible drop in resistivity at high fields often accompanied by destruction of the material Dielectric strengths of most insulating materials lie in the range from 10^4 10^7 volts/in at room temperature and low frequencies and decrease at higher temperatures The breakdown strengths of gases increase nearly linearly with pressure over a considerable range but at very low pressures the values also increase and the dielectric strength of a high vacuum is superior to gases at atmospheric pressure Dielectric breakdown is caused by an enormous increase in the number of charge carriers because of collisions or thermal ionization and field emission (*see* FIELD EMISSION)

Dielectric loss This is the power dissipated in a dielectric because of conduction processes This power loss results from thermal dissipation of the electrical energy expended by the field It is caused by molecular collisions It can be described by any of the following related parameters the conductivity σ the loss factor ϵ the power factor $\cos\theta$ and the loss tangent or dissipation factor $\tan\delta$ Of these only σ is applicable to direct current problems The conductivity σ is the current density I per unit field strength E in phase with the applied voltage The loss factor ϵ which is the imaginary part of the permittivity is related to the conductivity by $\sigma = \omega\epsilon''/\gamma$ where ω equals 2π times the frequency (*see* PERMITTIVITY) The power factor $\cos\theta$ is the ratio of conduction or loss current in phase with the applied voltage to the total current in any circuit and θ is the phase angle between current and voltage The dissipation factor $\tan\delta$ is the ratio of loss current to reactive or charging current where $\delta = 90° - \theta$ Expressed in terms of permittivity

$$\epsilon^* = \epsilon - i\epsilon \quad \text{and} \quad \cos\theta = \epsilon /|\epsilon^*|$$

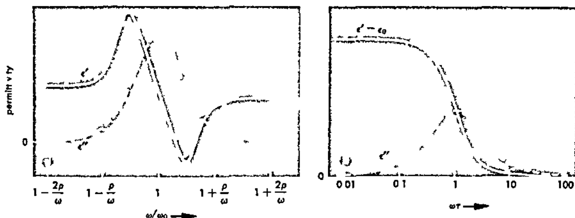


Fig 1 Dispersion of permittivity (a) Resonance spectrum ($\rho = f/m$) (b) Relaxation spectrum ($\tau = f/m\omega_0^2$)

where $|\epsilon^*| = \sqrt{\epsilon'^2 + \epsilon''^2}$

and $\tan \delta = \epsilon''/\epsilon'$ For low loss materials $\cos\theta$ and $\tan\delta$ are nearly equal

The power dissipated per unit volume $p = \sigma|\mathbf{E}|^2 = |\mathbf{I}||\mathbf{E}| \cos\theta$ This power loss increases at high temperatures and in many substances at high frequencies for a given field strength This effect is commercially employed in dielectric heating equipment for industrial and therapeutic purposes *See* DIELECTRIC HEATING

Dielectric constant The dielectric constant or permittivity relative to vacuum is important in many applications Materials with high dielectric constants are desirable for capacitors since they permit a reduction in size for a given capacitance while low dielectric constants are usually preferred for cable and transformer insulation For an extended discussion *see* DIELECTRIC CONSTANT *see also* CAPACITOR, INSULATION ELECTRIC

Dispersion The conductivity and dielectric constant or alternatively the permittivity have a frequency dependence determined by the molecular mechanisms of polarization in the dielectric Classically one may describe the action of an electric field $\mathbf{E}$ on a charged particle by the equation

$$m\frac{d^2\mathbf{r}}{dt^2} + f\frac{d\mathbf{r}}{dt} + k\mathbf{r} = e\mathbf{E}$$

where e m and $\mathbf{r}$ are the charge mass and position of the particle and f and k are the frictional and restoring force constants (The first term represents an acceleration and the remaining terms the net force acting on the particle) The solution for a sinusoidal field $\mathbf{E} = \mathbf{E}_0 \exp(i\omega t)$ is

$$\mathbf{r} = \frac{e\mathbf{E}}{m}\frac{1}{\omega_0^2 - \omega^2 + i\omega f/m}$$

where $\omega_0^2 = k/m$ The polarization

$$\mathbf{P} = (\epsilon^* - \epsilon_0)\mathbf{E}/\gamma = Ne\mathbf{r}$$

where N is the number of particles per unit volume and

in the cgs or mks systems respectively Thus

$$\epsilon^* = \epsilon_0 + \frac{\gamma Ne^2}{m}\frac{1}{\omega_0^2 - \omega^2 + i\omega f/m}$$

For $f/m < \omega_0$ the frequency dependence of ϵ^* is described as the resonance spectrum of a damped harmonic oscillator shown in Fig 1*a* For $f/m \gg \omega_0$, the frequency dependence approaches that of a relaxation circuit as shown in Fig 1*b*

The experimentally observed frequency dependences of the dielectric constant and permittivity can be satisfactorily accounted for in terms of resonance and relaxation processes Resonance dispersion is associated with changes in the electronic or vibrational energy of molecules the resonance frequencies usually occur at frequencies greater than 10^{12} cps Relaxation spectra occur for polar molecules in viscous media and for solids exhibiting interfacial polarization the time constant $\tau = f/m\omega_0^2$ is usually greater than 10^{-12} sec and may extend to very long periods *See* MOLECULAR STRUCTURE AND SPECTRA POLARIZATION DIELECTRIC

The presence of conduction phenomena characterized by a long relaxation time τ leads to the occurrence of dielectric hysteresis and absorption

Dielectric hysteresis is analogous to magnetic hysteresis (*see* HYSTERESIS MAGNETIC) It is the de

exp $(i\omega t)$ where $\omega > 1/\tau$ the polarization describes an ellipse when plotted versus $\mathbf{E}$ as shown in Fig 2*a* For $\omega \ll 1/\tau$ $\mathbf{P}$ approaches a single valued function of $\mathbf{E}$ given by $\mathbf{P} = \chi\epsilon_0\mathbf{E}$ where χ is the electric susceptibility Ferroelectric materials exhibit spontaneous polarization and show hysteresis even for nearly static fields (curve B Fig 2*b*) *See* FERROELECTRICS

Dielectric absorption or the dielectric aftereffect is the charging current or polarization which builds up or decays slowly when the field applied to a dielectric is changed It is usually caused by charge polarization and may in exceptional

Dielectric materials Vacuum or gaseous dielectrics other than air have had relatively little dielectric application except in electron tube devices voltage regulators and lightning arrestors

Dielectric liquids are principally employed for impregnating porous insulation in high voltage cables and capacitors and as insulating media for transformers and circuit breakers In the latter applications the heat transfer properties of the liquid are also important Mineral oils halogenated hydrocarbons and silicone oils are the most important commercial dielectric liquids

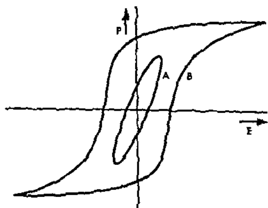


Fig 2 Dielectric hysteresis Curve A shows polarization for $\omega > 1/\tau$ Curve B shows quasi static polarization for a ferroelectric crystal

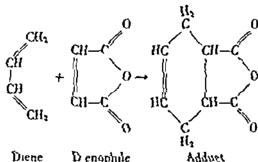
Solid dielectrics are employed for the vast majority of commercial applications Important solid dielectrics include many ceramics and glasses plastics and rubber minerals such as quartz mica magnesia and asbestos and paper and fibrous products The mechanical and thermal properties as well as the electrical response are important in the choice of a dielectric for a particular product For high mechanical strength and temperature resistance ceramic and mineral insulators are preferred while plastics and rubber are employed where flexibility is desired Low loss nonpolar dielectrics such as polyethylene or polystyrene are necessary for many ultra high frequency applications

Dielectric devices usually depend on nonlinear properties of dielectrics that is anisotropy of conductivity or polarization or saturation phenomena For some important applications of dielectrics see ANTENNA (AERIAL) WAVE GUIDE. [R D W]

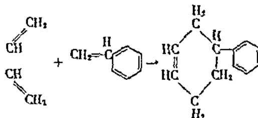
Bibliography A R von Hippel *Dielectrics and Waves* 1954 A R von Hippel *Dielectric Materials and Applications* 1954

Diels Alder reaction

The 1 4 addition of a conjugated diolefin to a compound known as a dienophile ...
or triple bon ...
conjugation ...
electron accep or ... The product is a six membered ring adduct



Diene Dienophile Adduct

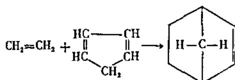


1 3 Butadiene Styrene 4 Phenyl 1 cyclohexene

Diene component With the exception of a few highly substituted diolefins which interfere with the reaction because of steric effects most alkyl and aryl homologs of butadiene react readily Alicyclic dienes are especially reactive except in those cases where the adduct formed would contain a bridged ring having a double bond at the bridgehead Aromatic compounds such as styrene in which part of the conjugation is in the ring react in a normal manner but heterocyclic dienes may show anomalous behavior Furan reacts normally to give a six membered ring containing a bridged oxygen whereas thiophene and pyrrole do not react in this manner Some highly substituted thiophenes react with dienophiles but in such cases the adducts are formed with elimination of H_2S

Dienophile component Most commonly dienophiles consist of compounds containing structure >C=C< or —C≡C— However dienophiles are not restricted to unsaturated carbon compounds and adducts form with >C=N— —C≡N —N=N— and —N=O as the dienophile component Dienophiles which have been studied in detail are indicated in the following list H C=CHR where R is H X CHO COOH COOMe COOEt COOCl CN C_6H_5 CH_2OH CH_2X CH_2CN CH_2COOH OCOMe COMe RHC=CHR where R is COOH COOMe COOEt OCOMe COMe RC≡CR where R is H COOH COOMe COOEt C_6H_5 COMe and H_2C=CR where R is COOEt CN COMe Quinones are especially reactive with dienes ...
... diene-
... nor
is the reaction retarded by the presence of oxidation inhibitors with which dienes are commonly treated to prevent formation of peroxides For reactive dienophiles such as maleic anhydride it is sufficient to mix the reactants in molar proportion usually in a solvent such as benzene and reaction

takes place at room temperature—often with the evolution of heat On the other hand the reaction of ethylene with cyclopentadiene to form bicyclo [2 2 1] 2 heptene requires temperatures in the range of 190–220°C and pressures in the range of 20–80 atm



Ethylene Cyclopentadiene Bicyclo[2 2 1] 2 heptene

Instances are known where the diene synthesis is a reversible reaction For example cyclohexene the adduct of butadiene and ethylene is quantitatively transformed into its components when subjected to high temperature for short contact time at low pressure Similarly, 4-vinylcyclohexene regenerates 2 moles of butadiene when heated

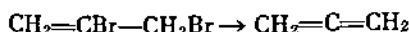
Industrially the diene synthesis is used in the production of the insecticides aldrin and dieldrin The adduct of butadiene and maleic anhydride is used in the synthesis of the important fungicide captan *See* ADDITION REACTION DIENE [C A C]

Bibliography K Alder *Newer Methods of Preparative Organic Chemistry* 1948

Diene

One of a class of organic compounds containing two ethylenic linkages (carbon to carbon double bonds) in the molecule They are sometimes termed diolefins or more simply dienes Structurally they may be classified as allenes isolated dienes or conjugated dienes (For structural representation *see* POLYALKENE) For a listing of the physical constants for some of the more common dienes *see* UNSATURATED HYDROCARBON

Allene is the simplest cumulative diene and it may be prepared readily by the dehalogenation of 2 3 dibromo 1 propene



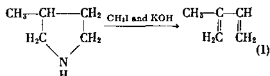
It hydrates readily in the presence of strong acid catalysts to give acetone and rearranges when heated with sodium to methylacetylene In those

chemistry

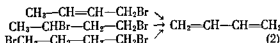
The ethylenic linkages in isolated dienes react as individual monoolefins except in those cases where polymerization may give rise to cyclic compounds

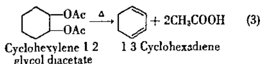
The most important group of compounds containing two double bonds are the conjugated dienes and unqualified literature references to dienes usually refer to conjugated dienes

Preparation of conjugated dienes Isoprene (2 methyl 1 3 butadiene) 1 3 butadiene and cyclopentadiene are produced on a tonnage basis from petroleum by catalytic dehydrogenation and partial pressure cracking processes In the laboratory conjugated dienes have been synthesized (1) by exhaustive methylation of cyclic imines and 1 4 diamines (2) by dehydrohalogenation of halogenated olefins or dihaloparaffins (3) by pyrolysis of esters of glycols and (4) by conversion of chlorohydrins to epoxides followed by catalytic dehydration Of special interest are the methods developed by Reppe in Germany for the synthesis of dienes from acetylene (5) Representative reactions illustrating these methods are as follows

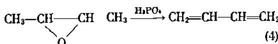


3 Methyl pyrrolidine Isoprene

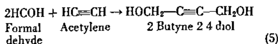


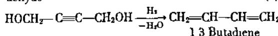


Cyclohexylene 1 2 glycol diacetate 1 3 Cyclohexadiene

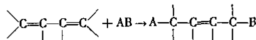


2 3 Butylene oxide Butadiene





Reactions of conjugated dienes Conjugated dienes are unique in that both ethylenic linkages in general react as a single unit in contradistinction to the cumulative or isolated dienes which can react in an independent manner This property is best illustrated by addition reactions in which the two components of a reagent add to the terminal carbons of the conjugated diene with the formation of a double bond at the site of the former single bond

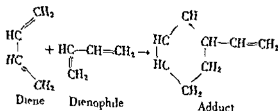


While 1 4 additions of this type (termed conjugate additions) are the most common for a conjugated diene they are not exclusive since many 1 2 additions are known On the other hand it may be noted that conjugated polyenes (containing three or more conjugated ethylene linkages) add almost exclusively in the 1 4 positions In any specific addition reaction the proportion of reactants

into the 1 2 or 1 4 positions may be greatly influenced by the presence of catalysts and the polarity of the solvent For example addition of chlorine to butadiene in the presence of ferric chloride yields 1 2 3 4 tetrachlorobutane addition of iodine on the other hand yields a mixture of the 1 4-diiodo 2 butene and 2 3 diiodo 1 butene Addition of bromine to butadiene in acetic acid solution yields predominantly 1 4 dibromo 2 butene and a smaller proportion of 3 4 dibromo 1 butene With hexane as solvent a reversal of the relative proportions of each of the isomers occurs Halogen acids add to butadiene in both the 1 2 and 1 4 positions to give a mixture of 3 halo 1 butene and 4 halo-2 butene

While most mono or polyolefins may be readily hydrogenated in the presence of catalysts such as nickel platinum black and copper chromite conjugated dienes may also be reduced to a mono olefin by chemical reagents such as zinc and acid or sodium and alcohol While catalytic reduction may add hydrogen in a 1 2 or 1 4 manner chemical reduction always proceeds exclusively in the 1 4 positions yielding the 2 3 olefin

a *cadmium* aluminum catalyst thiophene with hy*drogen sulfide* and *3* sulfolene with sulfur dioxide *The Diels Alder* reaction is a unique reaction of conjugated dienes involving 1 4 addition Where the diene acts as a dienophile it follows the general rule for Diels Alder syntheses and reacts across one terminal double bond



See CYCLOPENTADIENE DIELS ALDER REACTION

dioxide followed by acidification yields a dibasic acid of the dimerized diene "Isosebacic" ...

[C A C]

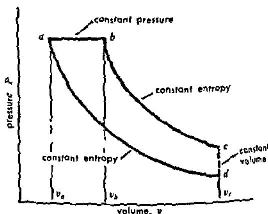
Diesel cycle

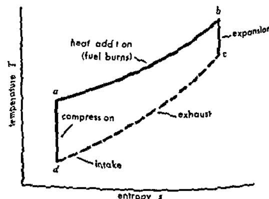
An internal combustion engine cycle in which the heat of compression ignites the fuel Compression ignition engines or diesel *engines are* thermodynamically similar to spark ignition engines The sequence of processes for both *types* is intake compression addition of *heat* expansion and exhaust Ignition and power control in the compression ignition engine are however, very different from those in the spark ignition engine

Usually a full unthrottled charge of air is drawn in during the intake stroke of a diesel engine A compression ratio between 12 and 20 is used in contrast to a ratio of 4 to 10 for the Otto spark ignition engine This high compression ratio of the diesel raises the temperature of the air during the compression stroke Just before top center on the compression stroke fuel is sprayed into the combustion chamber The high temperature of *the air* ignites the fuel which burns almost as soon as *it is* introduced adding heat The combustion products expand to produce power, and exhaust to complete the cycle

piston contains a unit air mass The metal cylinder head is alternately insulated and then uncovered for heat transfer

Air is compressed until the piston reaches the top of *the* stroke *Then* the air receives heat through *the cylinder head and expands at constant* pressure along path *a b* as shown in the diagram moving *the* piston part way down through the cylinder Then the cylinder head is insulated and the air completes its expansion along *path b-c* at constant entropy The cylinder head is uncovered and with the piston





Air standard diesel cycle

at the bottom of its stroke a constant volume heat rejection takes place on path c d The insulation is replaced and the cycle is completed with an i en tropic compression on path $d\,a$

An increase in compression ratio $r = v_a/v_c$ increases efficiency η the increase becoming less at higher compression ratios Another characteristic of the diesel cycle is the ratio of volumes at the end and at the start of the constant pressure heat addition process This cutoff ratio $r_c = v_b/v_c$ measures the interval during which fuel is injected For an engine to develop greater power output the cutoff ratio is increased and heat continues to be added further into the expansion stroke The air standard cycle shows that with less travel remaining during which to expend the additional heat energy as mechanical energy the efficiency of the engine is reduced Conversely efficiency increases as the cutoff ratio decreases so that a diesel engine is most efficient at light loads Specifically

$$\eta = 1 - \frac{1}{r^{k-1}}\left[\frac{r_c^k - 1}{k(r_c - 1)}\right]$$

where $k = c_p/c_v$ the ratio of specific heat of the working substance at constant pressure to its specific heat at constant volume In the limiting case when cutoff ratio r_c approaches unity diesel cycle efficiency approaches Otto cycle efficiency for cycles of the same compression ratio

In an actual engine with a given compression ratio the Otto engine has the higher efficiency However fuel requirements limit the Otto engine to a compression ratio of about 10 whereas a diesel engine can operate at a compression ratio of about 15 and consequently at a higher efficiency

In addition heat can be added earlier in the cycle by injecting fuel during the latter part of the compression process $d\,a$ This mode of operation is the dual combustion or semidiesel cycle With most of the heat added near peak compression semidiesel efficiency approaches Otto cycle efficiency at a given compression ratio [J B]

Bibliography L C Lichty *Internal Combustion Engines* 1951 E B Norris Eric Therkelsen and C E Trent *Applied Thermodynamics* 1955

Diesel engine

An internal combustion engine operating on a thermodynamic cycle in which the ratio of compression ($R_c = 15\pm$) of the air charge is sufficiently high to ignite the fuel subsequently injected into the combustion chamber (*see* DIESEL CYCLE) The engine differs essentially from the more prevalent mixture engine in which an explosive mixture of air and gas or air and the vapor of a volatile liquid fuel is made externally to the engine cylinder compressed to a point some 200 degrees below the ignition temperature and ignited at will as by an electric spark (*see* OTTO CYCLE) The diesel engine utilizes a wider variety of fuels with a higher thermal efficiency and consequent economic advantage under many service applications The true diesel engine as projected by R Diesel and as represented in most low speed engines such as about 300 rpm uses a fuel injection system where the injection rate is delayed and controlled to maintain constant pressure during combustion Adaptation of the injection principle to higher engine speeds such as 1000–2000 rpm, has necessitated departure from the constant pressure specification because the time available for fuel injection is so short (milliseconds) Combustion proceeds with little regard to the constant pressure specification High peak pressures may be developed Yet nonvolatile (distillate) fuels are burned to advantage in these engines which cannot be rigorously identified as true diesels but which properly should be called commercial diesels In ordinary parlance all such engines are classified as diesels

Identifying alternative features of diesel engine types include (1) two-cycle or four cycle [illegible] (2) horizontal or vertical piston [illegible] (3) single or multiple cylinder (4) large [illegible] hp) or small (50 hp) (5) cylinders in line opposed V or radial (6) single acting or double acting (7) high (1000–2000 rpm) low ([illegible] rpm) or medium speed (8) constant [illegible] variable speed (9) reversible or [illegible] (10) air injection or solid injection (11) [illegible] charged or unsupercharged and (12) [illegible] multiple fuel Section drawings of two [illegible] tive engines are given in Figs 1 and 2 [illegible] performance data are given in Table 1

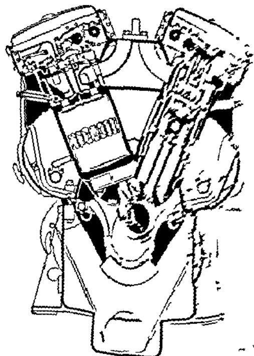


Fig 1 Section [illegible] General [illegible] Motive Division) [illegible] Knowlton ed [illegible] neers 9th ed [illegible] 195[illegible]

Table 1 Performance of selected diesel engine plants

Type of plant	Shaft horse power	Ratio of compression R_v	Brake mean effective pressure psi	Piston speed ft/min	Weight lb/in³ displacement	Weight lb/shp	Over-all thermal efficiency %
Air injection engine	300–5000	12–15	50–75	600–1000	3–8	25–200	30–35
Solid injection compression ignition							
Automotive	20–300	12–15	75–100	800–1800	2.5–4	7–25	25–30
Railroad	200–2500	12–15	60–90	800–1800	2.5–4	10–40	30–35
Stationary							
Unsupercharged	50–2500	12–15	70–80	600–1500	2.5–5	10–100	30–35
Supercharged	60–4000	10–13	110–125	600–1500	2.5–5	7.5–75	32–40
Dual fuel stationary							
Unsupercharged	50–2500	12–15	80–90	600–1500	2.5–5	10–100	30–35
Supercharged	60–4000	10–13	120–135	600–1500	2.5–5	7.5–75	32–40

Maximum diesel engine sizes (5000 kw) are less than steam turbines (500 000 kw) and hydraulic turbines (100 000 kw). They give high intrinsic and actual thermal efficiency (20–40%) with a sample comparative heat balance in Table 2. Variation in performance with load is shown in Fig. 3. Control of engine output is by regulation of the fuel supplied but without variation of the air supply (100% excess air at full load). *Supercharging* (10–15 psi) increases cylinder weight charge and consequently power output for a given cylinder size and engine speed. With two-cycle constructions scavenging air (approximately 5 psi) is delivered by crankcase compression, front end compression, or separate rotary reciprocating or centrifugal blowers. The engine cylinder may be without valves and with complete control of admission of scavenging air and release of spent gases in a two port construction, the piston covering and uncovering the ports, or the cylinder may have a single port (for admission or release) uncovered by the main piston at the outer end of its stroke and conventional cam operated valve in the cylinder head. The objective is to replace spent gases with fresh air by guided flow and high turbulence. The four cycle engine with its complement of admission and exhaust valves on each cylinder is most effective in scavenging. But the sacrifice of one power stroke out of every two is a frequent deterrent to its selection. Valves are exclusively of the poppet type with the burden of tightness and cooling dominant in the exhaust valve designs. *Cylinder heads* become complicated structures with valve porting, jacketing, and spray valve locations and the accommodation of them to effective combustion heat transfer and internal bursting pressures.

Table 2 Approximate allocation of losses in internal combustion engine plants

Type of loss	Mixture engines %	Injection (diesel) engines %
Output	20	33
Exhaust losses	40	33
Cooling system losses	40	33
Other	Less than 1	1
Total (input)	100	100

Distillate fuel (40° API, 19 000–19 500 Btu/lb, 135 000–140 000 Btu/gal) prevails with locomotive, truck, bus, and automotive applications. Lower speed engines (stationary and motorship service) burn heavier fuels (for example, 20° API, 18 500–19 000 Btu/lb, 145 000–150 000 Btu/gal). Alternative fuels are burned in dual fuel and gas diesel engines for stationary service. The main fuel is typically natural gas (90–95%) with oil (5–10%) used to control burning and to stabilize ignition. In the more prevalent liquid fuel injection system the technical problems are numerous and embrace such elements as pumps, spray nozzles, and combustion chambers for the delivery, atomization, and burning of the fuel in the hot compressed air. There must be accurate timing (measured in milliseconds) for the entire process to give clean complete combustion without undue excess air. Combustion characteristics of fuels are defined by rigorous specifications and include such factors as viscosity, flash point, pour point, ash, sulfur, basic sediment, water, Conradsen carbon number, cetane number, and diesel index.

Small size (<200 hp) engines are conveniently started by an electric motor and storage battery

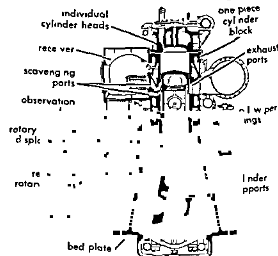


Fig. 2 Section through a Busch-Sulzer two-cycle diesel engine. (From A. E. Knowlton, ed., *Standard Handbook for Electrical Engineers, 9th ed.*, McGraw-Hill, 1957)

Larger engines use compressed air (about 200 psi) introduced through valves in the cylinder head Starting with engine-driven generator sets may be accomplished by motoring the generator

Cooling systems use water at 120–180°F with radiators cooling towers and cooling ponds employed for conservation and reclamation Lubrication costs can become prohibitive with inadequate

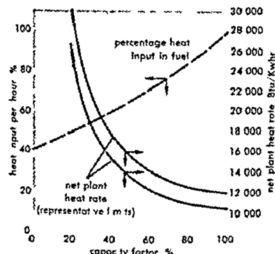


Fig 3 Heat rate and heat input curves for selected diesel-engine electric generating plants

engine maintenance Foundations must be designed to handle stress loadings and to reduce vibration Exhaust systems should be equipped with wave trap silencers or mufflers Filters on air and fuel supply are good insurance for engine reliability *See* INTERNAL COMBUSTION ENGINE [T B]

Bibliography ASME *Annual Oil Engine Power Cost Report* 1960 T Baumeister (ed) *Marks Mechanical Engineers Handbook* 6th ed 1958 Diesel Engine Manufacturers Association *Standard Practices* A W Judge *High Speed Diesel Engines* 5th ed 1957 L C Lichty *Internal Combustion Engines* 6th ed 1951

Diesel fuel

The diesel engine with a compression ratio as high as 15 1 ignites its fuel by the heat of compression of the air charge which reaches a temperature of nearly 1000°F It is not true that diesel engines can be operated successfully on waste oils of any kind The requirements as to cleanliness viscosity and the properties related to composition are severe Sulfur content must be kept low an increase in sulfur content from 0 2 to 1 0% accelerates wear and engine fouling quite perceptibly The fuels employed vary depending on the engine type from kerosine distillates to heavy oil residues For high speed automotive type engines a nonviscous easily atomized distillate boiling between 400 and 700°F is necessary Large low speed stationary or marine units however afford more time for introducing atomizing and burning the charge They successfully employ heavier less volatile oils

In gasoline engines the compression ratio is limited to relatively low values because of the occurrence of fuel knock In diesels the compression ratio must be high to bring about ignition and a fuel which ignites spontaneously at low temperatures has an advantage Gasoline is rated by octane numbers diesel fuel by an almost reciprocal scale called the cetane number scale In this scale cetane (*n* hexadecane $C_{16}H_{34}$) has a value of 100 and α methylnaphthalene $C_{11}H_{10}$ an aromatic hydrocarbon of high ignition temperature has a value of zero Most diesel fuels have cetane numbers of the order of 30 to 45 Additives like nitromethane can raise this value in much the same way that tetraethyllead raises the octane number of gasoline

It has been found that cetane number values can be approximated from certain physical properties of a half dozen methods proposed the best results have so far been obtained from the Calculated Cetane Index (ASTM D975 Appendix II) It is defined by an expression involving the log of the mid boiling point in degrees Fahrenheit (ASTM D86) and the API gravity *See* CETANE NUMBER DIESEL ENGINE OIL ANALYSIS [M SO]

Bibliography American Society for Testing Materials *Evaluation of Petroleum Products* 1944

Differential

A mechanism which permits a rear axle to turn corners with one wheel rolling faster than the other An automobile differential is located in the case carrying the rear axle drive gear (Fig 1)

The differential gears consist of the two side gears carrying the inner ends of the axle shafts meshing with two pinions mounted on a common pin located in the differential case The case carries a ring gear driven by a pinion at the end of the drive shaft This arrangement permits the drive to be carried to both wheels but at the same time as the outer wheel on a turn overruns the differential case the inner wheel lags by a like amount

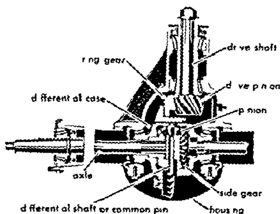


Fig 1 Commonly used rear-axle differential (Chrysler)

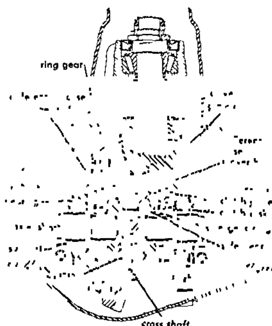


Fig 2 Nonspinning differential (*Salisbury Axle Division Dana Corp*)

Special differentials permit one wheel to drive the car by a predetermined amount even though the opposite wheel is on slippery pavement they have been used on racing cars for years and are now used by a number of car manufacturers (Fig 2)

In operation engine torque applied to the differential case causes the angular contacts on the case to bear on corresponding angles on the differential pinion pins Two contacts 180° apart push one pin to the right, the other two contacts 180° apart spaced 90° from the first push the opposite pin to the left On one of the pins are two pinions meshing with the right side gear, on the other are two pinions meshing with the left side gear Diameters on the pinions concentric with the teeth roll with the pinion thrust members and force the latter to apply clutches which connect the differential case directly to the pinion thrust members The pinions also load the side gears axially the latter contributing their thrusts to the pinion thrust members The drive thus passes from the differential case to the pinion thrust members splined to the respective rear axle shafts providing a frictional drive to one wheel even though the other is on ice or slippery pavement This nullifies the conventional differential action by a predetermined amount See TRANSMISSION AUTOMOTIVE

[F R M]

Bibliography F McFarland and E L Nash, Nonspinning differential gives increased traction *SAF J*, 65 19, 1957

Differential analyzer

A class of analog computers used for the solution of ordinary differential equations This class of machines which may be further subdivided into *mechanical*, electromechanical, and electronic differential analyzers is based upon operational or mathematical rather than direct, analogies See ANALOG COMPUTER

The technique for solving ordinary differential equations by analog means may be illustrated with the equation

$$\frac{dy}{dt} + ay = 0$$

to which the initial condition $y = 1$ when $t = 0$ is applied Solution of the original equation for dy/dt yields

$$\frac{dy}{dt} = -ay$$

If y is multiplied by $-a$, the result equals dy/dt the input that was required for the integrator The complete operation becomes a closed loop system (Fig 1) and the solution is generated automatically when the integrator is switched from the initial condition mode to the compute mode

Standard symbols. The drawing and interpretation of computer setup diagrams are facilitated by the use of standard symbols Although these symbols are not completely standardized the differences are relatively minor and usually cause no misunderstanding The symbols used in the remainder of this discussion are shown in the table

In terms of these symbols the setup diagram for the initial equation takes the form shown in Fig 2 Because the integrator provides an output of $-y$ rather than y for an input dy/dt, the sign of the initial condition must be reversed from that shown previously

The technique for solving a first order equation is readily extended for the solution of an nth order linear constant coefficient differential equation of the form

$$a_n \frac{d^n y}{dt^n} + a_{n-1} \frac{d^{n-1} y}{dt^{n-1}} + \quad \cdot \quad + a_1 \frac{dy}{dt} + a_0 y = f(t)$$

First the highest derivative is separated by putting the equation in the form

$$\frac{d^n y}{dt^n} = -\frac{1}{a_n}\left[a_{n-1}\frac{d^{n-1}y}{dt^{n-1}} + \qquad + a_1\frac{dy}{dt} + a_0 y - f(t)\right]$$

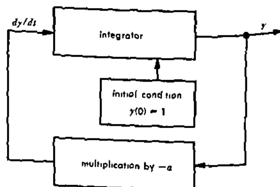
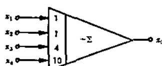
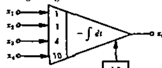


Fig 1 Block diagram representation of first-order differential equation

Analog computer symbols

Operation and symbol	Remarks
High gain inverting amplifier  $x_0 = -Ax_1$ where $\lvert A\rvert \to \infty$	The high gain amplifier represents the basic building block of the electronic differential analyzer
Summing amplifier  $x_0 = -(x_1 + x_2 + 4x_3 + 10x_4)$	The constants by which the various inputs are multiplied are typical of those normally provided in a summer or integrator
Summing integrator  $x_0 = -\int(x_1 + x_2 + 4x_3 + 10x_4)\,dt + \text{i c}$	Each of these units provides a sign reversal The initial condition is indicated in the box labeled i c
Coefficient multiplier  $x_0 = ax_1$ where $0 \leq a \leq 1$	The coefficient a is manually set before a solution is run
Generalized multiplier  $x_0 = x_1x_2/K$	This unit provides for multiplication of one dependent variable by another
Function generator  $x_0 = f(x_1)$	The abbreviation F G may be replaced by a simple graph of the function

Then the availability of this highest derivative is assumed and it is integrated n times to yield y The various derivatives are multiplied by the appropriate coefficients summed added to $f(t)$ and finally multiplied by $-1/a_n$ to give the highest derivative d^ny/dt^n Because the signs of outputs of successive integrators alternate care must be taken to see that each term is added with the correct sign This technique may be illustrated by reference to the setup diagram of Fig 3 for solving the third order equation

$$a_3\frac{d^3y}{dt^3} + a_2\frac{d^2y}{dt^2} + a_1\frac{dy}{dt} + a_0y = f(t)$$

In this setup the assumption has been made that all the coefficients in the equation are positive and less than unity The occurrence of negative coefficients would require addition or removal of inverting amplifiers and coefficients larger than unity would require the insertion of amplifiers with greater than unity The problems involved in arriving at a practical computer setup can be appreciated more readily after discussion of scale factors

Scale factors After the basic block diagram for the representation of a physical system has been determined scale factors must be assigned that relate (1) the amplitudes of variables within the computer to the magnitudes of the corresponding mathematical variables in the differential equation to be solved and (2) the time required for an event to take place in the machine (real time) to the time required for it to occur in the problem being investigated (problem time)

In an electronic computer the relationship between an equation variable y_1 and the corresponding computer voltage v_1 can be written

$$v_1 = a_1y_1$$

In general the scale factor a_1 is a dimensional con-

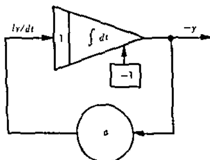


Fig 2 Computer representation of first order homogeneous differential equation

stant because y_1 and v_1 usually have different dimensions For example if y_1 is a distance measured in feet and 5 volts of v_1 correspond to 1 ft of y_1 the relationship becomes

$$v_1 = \frac{5 \text{ volts}}{1 \text{ ft}} y_1$$

The scale factors used in a computer are not arbitrary but are based upon the physical characteristics of the computer components Physical variables in each component are limited to certain maximum and minimum values In an electronic differential analyzer the output voltage of each computing component usually must remain between -100 and $+100$ volts Furthermore the rates of change of computer variables are limited if electromechanical components are used On the other hand because of the presence of noise large inaccuracies result if the computer variables are restricted to extremely small values Consequently the amplitude of the computer variables should be made as large as possible without exceeding the limiting values

Within these general limitations scale factors associated with linear operations may be selected arbitrarily provided that all inputs to a single summing amplifier have the same scale factor Although exactly the same principles are used in determining the scale factors for nonlinear operations much less freedom exists when nonlinear operations are to be performed

Machine time τ and problem time t are related by the coefficient a_t called the time scale factor Where only a portion of a problem is represented with computer equipment and a person or some physical equipment from the system being studied is also involved the computer must operate with a time scale factor of unity if meaningful results are to be achieved For example the computer portion of a flight trainer used for training aircraft pilots must operate with a unity time scale factor or as is frequently stated must operate in real time if the pilot is to experience realistic responses when he actuates the controls of the simulated aircraft On the other hand a computer used in studying an electron ballistics problem would operate with a large time scale factor or a large time scale extension

A change in the time scale on which a computer is operating can be effected by changing only those components performing operations inherently dependent on time For example the solution time for a third order linear differential equation with the setup described could be doubled merely by halving the gains of each of the three integrators No change in the initial conditions the summing circuit or the coefficient potentiometers would be required

As a practical matter it is desirable to arrange an analog computer (unless it is of the high speed repetitive type) for a solution time in the range of 30 sec to 2 min A lower limit is set by the speed of response of mechanical elements such as servo units or recorders while an upper limit is set by integrator drift [W W S]

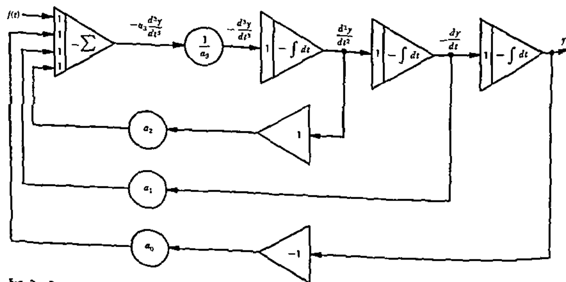


Fig 3 Setup for solution of linear third-order differential equation

Differential equation

Differential equations are divided into ordinary differential equations and partial differential equations, according to the number of independent variables involved

Classification. An ordinary differential equation is an equation of the general form

$$f(x,u,u',u'', \quad ,u^{(n)}) = 0$$

that is, a relationship between the single independent variable x, a function $u(x)$ of the single independent variable x, and the successive derivatives

$$u' = \frac{du}{dx}, u'' = \frac{d^2u}{dx^2}, \quad , u^{(n)} = \frac{d^nu}{dx^n}$$

of the dependent variable $u(x)$ The function $f = f(x,u_0,u_1 u_n, \quad ,u_n)$ is a given function of the $n+2$ independent variables x, u_0, u_1, $, u_n$ *The positive integer n is referred to as the order of* the differential equation, n is the order of the highest derivative of $u(x)$ which actually appears in the differential equation The differential equation is said to be linear or nonlinear according to whether the function $f(x,u_0,u_1,u_2, . \quad ,u_n)$ of the $n+2$ independent variables x, u_0, u_1, u_2, $, u_n$ is a linear or nonlinear function of the last $n+1$ variables u_0, u_1, u_2, $, u_n$

The differential equation $u' + x^2u = 0$ is a linear first order differential equation Here

$$f(x,u_0 u_1) = u_1 + x^2u_0$$

is a function of the three independent variables x, u_0, u_1 which is linear in u_0, u_1

The differential equation $u'' + u' + u^2 = 0$ is a nonlinear second order differential equation Here $f(x u_0,u_1,u_2) = u_2 + u_1 + u_0^2$ is a function of the four independent variables x, u_0, u_1, u_2 which is nonlinear in u_0 u_1, u_2 A solution of an ordinary differential equation $f(x u,u',u'', \quad ,u^{(n)}) = 0$ is a function $v(x)$ which satisfies the equation identically on some interval $a < x < b$, that is such that $f[x,v(x),v'(x),v''(x), \quad ,v^{(n)}(x)] = 0$ for all x on $a < x < b$

A partial differential equation in two independent variables is an equation of the general form

$$f\left(x\ y,u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial x\, \partial y}, \frac{\partial^2 u}{\partial y^2}, \quad , \frac{\partial^{i+j} u}{\partial x^i\, \partial y^j}, \quad , \frac{\partial^{m+n} u}{\partial x^m\, \partial y^n}\right) = 0$$

that is a relationship between the two independent variables x and y, a function $u(x,y)$ of the two independent variables x and y, and the successive partial derivatives

$$\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \quad , \frac{\partial^{i+j} u}{\partial x^i\, \partial y^j}, \quad , \frac{\partial^{m+n} u}{\partial x^m\, \partial y^n}$$

of the function $u(x,y)$, up to the order $i \leq m$ in the variable x and up to the order $j \leq n$ in the variable y The function

$$f(x,y,u_{00},u_{10},u_{01},u_{20},u_{11},u_{02}, \quad ,u_{ij}, \quad ,u_{mn})$$

is a given function of the independent variables indicated The positive integer $m+n$ is referred to as the order of the partial differential equation, $m+n$ is the order of the "highest" partial derivative of the function $u(x,y)$ which actually appears in the equation The differential equation is said to be linear or nonlinear according to whether the function f is a linear or nonlinear function of the last mentioned variables u_{00}, u_{10}, u_{01}, u_{20}, u_{11}, u_{02}, $, u_{mn}$

The differential equation $u_x + u_y = 0$ is a linear first order partial differential equation in the two independent variables x and y Here

$$u_x = \frac{\partial u}{\partial x} \qquad u_y = \frac{\partial u}{\partial y}$$

and $f(x,y,u_{00}\ u_{10},u_{01}) = u_{10} + u_{01}$ is a linear function of the independent variables u_{00}, u_{10}, u_{01}

The differential equation $u_{xx} + 3u_{xy} + u_{yy} = 0$ is a linear second order partial differential equation in the two independent variables x and y Here

$$u_{xx} = \frac{\partial^2 u}{\partial x^2} \qquad u_{xy} = \frac{\partial^2 u}{\partial x\, \partial y} \qquad u_{yy} = \frac{\partial^2 u}{\partial y^2}$$

and $f(x,y\ u_{00},u_{10}\ u_{01}\ u_{20},u_{11}\ u_{02}) = u_{20} + 3u_{11} + u_{02}$ is a linear function of the independent variables u_{00}, u_{10}, u_{01}, u_{20}, u_{11} u_{02}

The differential equation $u_x^2 + u_y^2 - u = 0$ is a nonlinear first order partial differential equation in the two independent variables x and y Here

$$f(x,y\ u_{00}\ u_{10},u_{01}) = u_{10}^2 + u_{01}^2 - u_{00}$$

A solution of a partial differential equation

$$f\left(x,y,u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \quad , \frac{\partial^{m+n} u}{\partial x^m\, \partial y^n}\right) = 0$$

is a function $v(x,y)$ which satisfies the equation identically on at least some rectangle $a < x < b$, $c < y < d$ that is such that

$$f\left[x,y,v(x,y)\ \frac{\partial v}{\partial x}(x,y), \frac{\partial v}{\partial y}(x,y), \quad , \frac{\partial^{m+n} v}{\partial x^m\, \partial y^n}(x,y)\right] = 0$$

for all (x,y) on $a < x < b, c < y < d$

There are also partial differential equations in more than two independent variables for example,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

is a linear second order partial differential equation in the three independent variables x y, and z Besides, so far, mention has been made only of single differential equations but systems of differential equations also occur For example

$$\frac{du}{dx} = v^2 \qquad \frac{d^2v}{dx^2} = u$$

is a nonlinear second-order system of two ordi

differential equations in the two functions $u(x)$ and $v(x)$. The system of differential equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

is a linear first order system of partial differential equations in the two functions $u(x, y)$ and $v(x, y)$. The independent variables and the functions involved may be in general real or complex valued.

Boundary value problems. The mathematical analysis of many physical problems leads to a consideration of boundary value problems for differential equations. The physical quantity of interest is found to be represented by a function (or functions) which satisfy a differential equation (or a system of differential equations). Besides the differential equations the sought functions are also required to satisfy certain other conditions which will be referred to collectively as boundary conditions. Thus a boundary value problem consists of a differential equation (or a system) plus a set of boundary conditions to be satisfied by an a priori unknown function (or set of functions). For example, the determination of the steady state temperature in a three dimensional unit cube whose surface is maintained at a given temperature has as its mathematical counterpart the boundary value problem consisting of the determination of a real valued function $u(x, y, z)$ satisfying the partial differential equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

in the interior of the cube $0 < x < 1$, $0 < y < 1$, $0 < z < 1$, and the boundary condition $u(x, y, z) = g(x, y, z)$ on the surface of the cube, where g is a given function. The remainder of this article will be devoted to a discussion of several important types of ordinary and partial differential equations of the first and second orders and of certain boundary value problems connected with them. Both the independent and the dependent variables will be assumed to be real without further explicit mention.

Ordinary differential equations. A typical problem for the first order differential equation

$$\frac{dy}{dx} = f(x, y)$$

in which the dependent variable will be written $y(x)$ to follow a customary notation, consists of the determination of a solution $y(x)$ which has an assigned value y_0 at $x = x_0$. A basic existence and uniqueness result for this (Cauchy) problem is that if the given function $f(x, y)$ of the two independent variables (x, y) has continuous first partial derivatives

$$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$$

of the first order in the rectangle

$$x_0 - a < x < x_0 + a, \quad y_0 - b < y < y_0 + b$$

where $a > 0$, $b > 0$ and $|f(x, y)| \leq M$ on the same rectangle, where $M > 0$ is a constant, then there is one and only one function $y(x)$, continuously differentiable on the interval

$$|x - x_0| < \min\left(a, \frac{b}{M}\right)$$

which satisfies the differential equation on this interval and is such that $y(x_0) = y_0$. This is purely an existence and uniqueness result which is applicable in the general case of the equation but which does not give the solution explicitly. For many special forms of the equation the determination of the general solution may be carried out by well known methods. In order to illustrate several of these methods it is convenient to rewrite the equation in differential form $M(x, y)\, dx + N(x, y)\, dy = 0$, where $f(x, y) = -M(x, y)/N(x, y)$. The following special cases of this form are readily integrable.

Variables separable. In this case both functions M and N are of separated form: $M(x, y) = F(x)G(y)$ and $N(x, y) = H(x)I(y)$. Example:

$$\frac{dy}{dx} = x/y$$

may be rewritten $y\, dy - x\, dx = 0$, that is

$$d\left(\frac{y^2 - x^2}{2}\right) = 0$$

and its general solution is $y^2 - x^2 = c$, with c an arbitrary constant.

Homogeneous functions. Here M and N are homogeneous functions of the same degree n, that is

$$M(\lambda x, \lambda y) = \lambda^n M(x, y)$$
and
$$N(\lambda x, \lambda y) = \lambda^n N(x, y)$$

Putting $y = vx$, where v is a new dependent variable, leads to a separable equation in v and x. In the example just mentioned the equation is homogeneous of degree one.

Exact functions. If there is a function $v(x, y)$ such that its differential $dv = v_x dx + v_y dy$ equals $M\, dx + N\, dy$, then the differential equation $M\, dx + N\, dy = 0$ is said to be exact. See the example above where

$$v(x,y) = \frac{y^2 - x^2}{2}$$

Integrating factors. If the equation $M\, dx + N\, dy = 0$ is not exact, it may be possible to multiply it by an (integrating) factor $u(x, y)$ so as to make it exact. Example:

$$\frac{dy}{x} - \frac{dx}{y} = 0$$

is rendered exact (see example above) by multiplication by $u(x, y) = xy$. The linear equation

$$\frac{dy}{dx} + p(x)y = q(x)$$

can be readily integrated since multiplication by

$\exp \int - p\, dx$ makes the left hand side equal to

$$\frac{d}{dx}\left[y(x) \exp \int - p\, dx\right]$$

Ordinary linear second-order equations A typical problem for the linear second order differential equation

$$\frac{d^2y}{dx^2} + a_1(x)\frac{dy}{dx} + a_0(x)y = 0$$

consists of the determination of a solution $y(x)$ which satisfies the initial conditions $y(x_0) = y_0$ $y'(x_0) = y'_0$ where y_0 y'_0 are given numbers If the coefficients $a_1(x)$ $a_0(x)$ are constants then the general solution of the equation reduces to the algebraic problem of finding the roots of a quadratic equation Substituting $y = e^{mx}$ with m a constant in the differential equation gives the quadratic equation $m^2 + a_1 m + a_0 = 0$ for the determination of the solutions of the differential equation of the form e^{mx} For example putting $y = e^{mx}$ in the equation $y'' - y = 0$ gives $m^2 - 1$ that is $m = \pm 1$ and hence the general solution of the differential equation is $y = c_1 e^x + c_2 e^{-x}$ where c_1 and c_2 are arbitrary constants Many of the ordinary differential equations occurring in mathematical physics fall under the general category of linear second order ordinary differential equations

Bessel's equation One of the solutions of Bessel s equation

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - m^2)y = 0$$

is the Bessel function of order n

$$J_n(x) = \sum_{k=0}^{\infty} \frac{(-1)^k (x/2)^{m+2k}}{\Gamma(k+1)\Gamma(m+k+1)}$$

where Γ denotes Euler s gamma function

Legendre s equation One of the solutions of the Legendre equation

$$(1 - x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + n(n+1)y - 0$$

when n is a nonnegative integer is

$$P_n(x) - \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2 - 1)^n$$

the Legendre polynomial

Gauss hypergeometric equation If c is not zero or a negative integer one of the solutions of Gauss equation

$$x(1 - x)\frac{d^2y}{dx^2} + [c - (a + b + 1)x]\frac{dy}{dx} - aby = 0$$

for $-1 < x < 1$ is the hypergeometric series

$F(a\ b\ c,x)$

$$= 1 + \frac{a\ \ b}{1\ \ c}x + \frac{a(a+1)\ \ b(b+1)}{1\ \ 2\ \ c(c+1)}x^2 + \ldots$$

Partial differential equations A typical problem for the first-order partial differential equation in two independent variables x y

$$F[x\ y, z(x\ y)\ p(x\ y)\ q(x\ y)] - 0$$

where

$$p = \frac{\partial}{\partial x} \qquad q = \frac{\partial}{\partial y}$$

is the initial value or Cauchy problem This consists of the determination of a solution $z(x\ y)$ of the equation which passes through a given curve in $(x\ y, z)$ space That is if $x = f(s)$ $y = g(s)$ $z = h(s)$ are the equations of the given curve in parametric form the requirement on the solution $z(x\ y)$ is that $h(s) = z[f(s)\ g(s)]$ The solution of this problem can be reduced to the integration of a system of five ordinary differential equations (called the characteristic equations of the first order partial differential equation) for five functions $x(t)$ $y(t)$ $z(t)$ $p(t)$ $q(t)$ namely

$$\frac{dx}{dt} - \frac{\partial F}{\partial p}(x\ y\ \ p\ q) \qquad \frac{dy}{dt} - \frac{\partial F}{\partial q}(x\ y, z, p\ q)$$

$$\frac{d}{dt} = \frac{\partial F}{\partial p}\ p + \frac{\partial F}{\partial q}\ q \qquad \frac{dp}{dt} = -\left(\frac{\partial F}{\partial x} + p\frac{\partial F}{\partial z}\right)$$

$$\frac{dq}{dt} - -\left(\frac{\partial F}{\partial y} + q\frac{\partial F}{\partial}\right)$$

The general linear partial differential equation of the second order in the two independent real variables x and y is

$$a\frac{\partial^2 u}{\partial x^2} + 2b\frac{\partial^2 u}{\partial x\, \partial y} + c\frac{\partial^2 u}{\partial y^2} + d\frac{\partial u}{\partial x} + e\frac{\partial u}{\partial y} + fu - g$$

where the coefficients a b g are real valued functions of x and y The equation is called homogeneous (nonhomogeneous) according to whether g is (is not) identically zero At any given point $(x\ y)$ at which at lea t one of the coefficients a b c is not zero an equation may be classified as one of three types *elliptic parabolic or hyperbolic* according to whether b^2 ac is less than equal to or greater than zero respectively at the point Equations of mixed type (whose type varies from point to point) are of special importance in fluid dynamics Equations with constant coefficients have the same type at every point.

Elliptic type This includes Laplace s equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

here $a = c = 1$ $b = 0$ A typical boundary value problem such as the determination of the steady state temperature $u(x\ y)$ in a semi infinite plate $y > 0$ $-\infty < x < +\infty$ when the edge $y = 0$ $-\infty < x < +\infty$ is held at a fixed temperature $f(x)$ consists of finding $u(x\ y)$ satisfy tion for $y > 0$ $-\infty < x < +\infty$ $f(x)$ for $-\infty < x < +\infty$ where function

Parabolic type This includes Fourier s heat equation

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial u}{\partial y} = 0$$

here $a = 1$ $b = c = 0$ and y denotes the time A typical boundary value problem [the determination of the transient temperature $u(x\ y)$ in an infinite wire when its initial temperature $u(x\ 0)$ is prescribed] consists of finding $u(x\ y)$ satisfying the equation for $y > 0$ $\infty < x < +\infty$ while $u(x\ 0) = f(x)$ for $-\infty < x < +\infty$ where $f(x)$ is a given function

Hyperbolic type This is illustrated by d Alembert s wave equation

$$\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} \quad 0$$

here $a = -c - 1$ b 0 and y *again denotes the* time A typical boundary value problem such as the determination of the transient displacement $u(x\ y)$ in an infinite string when the initial displacement $u(x\ 0)$ and the initial normal velocity

$$\frac{\partial u}{\partial y}(x\ 0)$$

are prescribed consi ts of the determination of a solution $u(x\ y)$ of the equation for

$$y > 0 \qquad -\infty < x < +\infty$$

satisfying $u(x\ 0) = f(x)$ and

$$\frac{\partial u}{\partial y}(x\ 0) = g(x)$$

for $-\infty < x < +\infty$ where $f(x)$ and $g(x)$ are *given functions* *See* BESSEL FUNCTIONS CALCULUS DIFFERENTIAL AND INTEGRAL DIFFERENTIATION

[J B D]

Differentiation

A mathematical operation performed on a function *to determine the effect of* a change in the value of

...

definition ... at x_0 is by

$$f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

For generalities about derivatives and calculus *see* CALCULUS DIFFERENTIAL AND INTEGRAL *In the* quotient on the right in the definition of $f'(x_0)$ x is restricted to the interval on which f *is defined* *If* $y = f(x)$ $f'(x_0)$ is also denoted by

$$\left(\frac{dy}{dx}\right)_{x=x_0}$$

The l mit defining $f'(x_0)$ may not exist *If it does* f is called differentiable at x_0

The derivative of $f'(x)$, called the second derivative is denoted by

$$f''(x) \quad \text{or} \quad \frac{d^2y}{dx^2}$$

A function f is called continuous at x_0 if x_0 is in the domain of f and

$$\lim_{x \to x_0} f(x) = f(x_0)$$

The precise formulation of the limit concept used here is the following Let g be a function and let A be a number Then

$$\lim_{x \to x_0} g(x) = A$$

means that to each positive number ϵ corresponds some positive number δ such that $|g(x) - A| < \epsilon$ whenever x is a number in the domain of g such that $x \neq x_0$ and $|x - x_0| < \delta$ It is required of g that its domain shall contain numbers x as close to x_0 as desired but different from x_0 The domain of g may also contain x_0 but this is irrelevant

It is a theorem that if f is differentiable at x_0, then f is continuous at x_0 However a function can be continuous at x_0 but not differentiable there An example is $f(x) = |x|$ at $x_0 = 0$ It is even possible for a function to be continuous on an interval and yet not differentiable at any point of this interval

The chief elementary applications of d fferentiation are (1) in expressing rates of change (velocity acceleration) and in solving problems where through functional relationship the rate of change of one variable is calculated when the rate of change of another variable is known (2) in studying graphs of functions and more generally in studying curves in the plane or in space of three dimensions (3) in expressing scientific laws or principles in the form of differential equations (4) in the expression of various extensions and applications of the law of the mean including such topics as l Hospital s rule and Taylor s formula or series

Principles The general technique of differentiation is built upon the rules for differentiating combinations of differentiable functions If $u = f(x)$ and $v = g(x)$ are functions differentiable for the same values of x then $u + v$ and uv are differentiable and so is u/v if $v \neq 0$ The formulas are

$$\frac{d}{dx}(u + v) = \frac{du}{dx} + \frac{dv}{dx} \qquad \frac{d}{dx}(uv) = u\frac{dv}{dx} + v\frac{du}{dx}$$

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v\dfrac{du}{dx} - u\dfrac{dv}{dx}}{v^2}$$

If u is constant in value then $du/dx = 0$ A very powerful instrument of technique is furnished in the chain rule for composite functions If y is a differentiable function of u and u is a differen

tiable function of x, then y is a differentiable function of x, and

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

In functional notation if $y = f(x)$ and $u = h(x)$, then $y = F(x)$, where $F(x) = f[h(x)]$, and then $F'(x) = f'[h(x)]h'(x)$

The technique of differentiation also leans upon facts about inverse functions If $y = f(x)$, where f is a differentiable function and $f'(x)$ is either always positive or always negative on an interval, for example, when $a \leqq x \leqq b$, then for each y from $f(a)$ to $f(b)$ inclusive there is just one x such that $a \leqq x \leqq b$ and $y = f(x)$ Thus there is defined a function $x = g(y)$ such that $x = g(y)$ is equivalent to $y = f(x)$ with x and y restricted as indicated This function g is differentiable, and

$$g'(y) = \frac{1}{f'(x)} = \frac{1}{f'[g(y)]}$$

An example: $y = \sin x$, $x = \sin^{-1} y$ (sometimes written $x = \arcsin y$), where

$$-\frac{\pi}{2} \leqq x \leqq \frac{\pi}{2} \quad \text{and} \quad -1 \leqq y \leqq 1$$

Here $f'(x) = \cos x$,

$$\frac{d}{dy}(\sin^{-1} y) = \frac{1}{\cos x} = \frac{1}{\sqrt{1 - \sin^2 x}} = \frac{1}{\sqrt{1 - y^2}}$$

Algebraic and transcendental functions The functions studied in elementary calculus are of two kinds algebraic and transcendental The basic differentiation formula for algebraic functions is

$$\frac{dy}{dx} = nx^{n-1} \quad \text{if} \quad y = x^n$$

where n is any rational number This rule may be combined with the chain rule and used in connection with the rules for dealing with sums products and quotients The differentiation of algebraic functions in general may require use of implicit function theorems For a discussion of these theorems *see* PARTIAL DIFFERENTIATION The elementary transcendental functions are the trigonometric functions, the logarithm functions, and their inverses

Trigonometric functions In calculus the trigonometric functions are defined on the assumption that angles are measured in radians so that $\sin x$ means the sine of x radians Differentiation of $f(x) = \sin x$ is based on the fact that

$$\frac{\sin x}{x} \to 1 \quad \text{as} \quad x \to 0$$

This is equivalent to $f'(0) = 1$ By combining this result with trigonometric identities, the two basic formulas

$$\frac{d}{dx}\sin x = \cos x \qquad \frac{d}{dx}\cos x = -\sin x$$

are derived Then the derivatives of the other functions are worked out by using the rule for quotients The results are

$$\frac{d}{dx}\tan x = \sec^2 x \qquad \frac{d}{dx}\cot x = -\csc^2 x$$

$$\frac{d}{dx}\sec x = \sec x \tan x \qquad \frac{d}{dx}\csc x = -\csc x \cot x$$

These formulas may also be combined with the chain rule For example,

$$\frac{d}{dx}\sin u = \cos u \frac{du}{dx}$$

The inverse trigonometric functions are differentiated by the methods for inverse functions, as explained earlier The definitions are

$y = \sin^{-1} x$ means $x = \sin y$ and $-\frac{\pi}{2} \leqq y \leqq \frac{\pi}{2}$

$y = \cos^{-1} x$ means $x = \cos y$ and $0 \leqq y \leqq \pi$

$y = \tan^{-1} x$ means $x = \tan y$ and $-\frac{\pi}{2} < y < \frac{\pi}{2}$

$y = \cot^{-1} x$ means $x = \cot y$ and $0 < y < \pi$

The differentiation formulas are

$$\frac{d}{dx}\sin^{-1} x = \frac{1}{\sqrt{1 - x^2}} \qquad \frac{d}{dx}\cos^{-1} x = \frac{-1}{\sqrt{1 - x^2}}$$

$$\frac{d}{dx}\tan^{-1} x = \frac{1}{1 + x^2} \qquad \frac{d}{dx}\text{ctn}^{-1} x = \frac{-1}{1 + x^2}$$

The inverses of the secant and cosecant functions are little used and there is no standard usage about the definitions needed to make them single valued

Exponentials and logarithms These two functions go together a logarithm function being the inverse of an exponential function or vice versa The traditional treatment of these functions in differential calculus was for many years as follows The nature of exponentials was assumed as known from algebra and the definition $y = \log_a x$ if $x = a^y$ (where $a > 0$, $a \neq 1$, and $x > 0$) as well as the algebraic properties of logarithms was also assumed as known As a first step toward the differentiation of a logarithm function it was traditional to show that $(1 + t)^{1/t}$ approaches a limit denoted by e as t approaches 0 Moreover, $2 < e < 3$, and $e = 2\,718$ approximately It can then be shown that

$$\frac{d}{dx}\log_a x = \frac{1}{x}\log_a e$$

This formula suggests the advantage of choosing $a = e$ as the base for logarithms The base e logarithm of x, called the natural logarithm of x, is denoted by $\log x$ with no subscript, or by $\ln x$ (The latter notation is favored by engineers and many physical scientists) Then $y = \ln x$ is equivalent to $x = e^y$, and $y = e^x$ is equivalent to $x = \ln y$ The derivative of e^x can be worked

by the rule for inverse functions The simple formulas are

$$\frac{d}{dx}\ln x = \frac{1}{x} \quad \text{and} \quad \frac{d}{dx}e^x = e^x$$

For an arbitrary base a where $a \neq 1$ and $0 < a$ the formulas are

$$\frac{d}{dx}\log_a x = \frac{\log_a e}{x} \quad \text{and} \quad \frac{d}{dx}a^x = \log_e a$$

It is often convenient to know that

$$\log_a x = (\log_a b)(\log_b x)$$

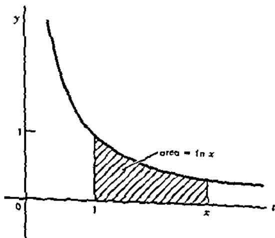
In particular with $a = 1$ one obtains

$$\log b = (\log_b a)^{-1}$$

The modern trend is to recognize that neither exponentials nor logarithms are adequately defined and studied by students before they come to calculus In any case satisfactory discussion of these functions requires the theory of limits Hence it is reasonable to use calculus itself to define logarithms and develop their properties When this is done the development proceeds as follows The natural logarithm function is defined when $x > 0$ by

$$\ln x = \int_1^x \frac{dt}{t}$$

This means that $\ln x$ is the area between the curve $y = 1/t$ and the t axis from $t = 1$ to $t = x$ reckoned as positive if $x > 1$ and reckoned as negative if $0 < x < 1$ (see the illustration)



The natural logarithm function

By use of the fundamental properties of integrals and the relation between differentiation and integration it can be shown quite directly that

$$\ln 1 = 0 \quad \frac{d}{dx}\ln x = \frac{1}{x}$$

and that $\ln (AB) = \ln A + \ln B$ Moreover $\ln x$ is a continuous function of x which increases as x increases and is such that $\ln x \to +\infty$ as $x \to +\infty$ and $\ln x \to -\infty$ as $x \to 0$ There is then a well defined inverse function (call it the E function) such that $y = \ln x$ is equivalent to $x = E(y)$ if $x > 0$ The number $E(1)$ is denoted by e, that is e is the unique positive number such that $\ln e = 1$ The rules about inverse functions show that $E(x) = F(x)$ If $a > 0$ and x is an integer it is possible to show easily that

$$a^x = E(x \ln a)$$

This formula serves as the general definition of a^x when x is not an integer, and the results are consistent with what is expected so that the usual exponent laws are valid The general definition of logarithms from this point of view is

$$\log_a x = \frac{\ln x}{\ln a}$$

In particular $\log_e x = \ln x$ because $\ln e = 1$

In this method of developing properties of exponentials and logarithms by calculus the fact that $(1 + t)^{1/t} \to e$ as $t \to 0$ is not needed but it can be proved more easily than in the older traditional development after everything else has been worked out

The number e enters naturally into the concept of continuous compounding of interest If the sum $\$P$ is placed at interest at the nominal rate of $100r\%$ compounded n times a year the accumulated sum $\$S$ after t years is

$$S = P\left(1 + \frac{r}{n}\right)^{nt}$$

If now n is increased indefinitely, so that interest is compounded more and more frequently the fact that

$$\lim_{n\to\infty}\left(1 + \frac{r}{n}\right)^{n/r} = e$$

shows that in the limit of continuously compounded interest the formula for the accumulated sum after t years is $S = Pe^{rt}$ This formula is characterized by the differential equation

$$\frac{dS}{dt} = rS$$

which expresses what is sometimes called the *law of natural growth* In general if y depends on x in such a way that $dy/dx = ky$ where k is a nonzero constant then $y = y_0e^{kx}$ where $y = y_0$ when $x = 0$ This situation occurs in radioactive decay and in many types of growth and diminution processes in chemistry and natural science

Applications Several of the important applications of differentiation are discussed in this section

Velocity and acceleration If a point moves on the x axis with coordinate x at time t its velocity is dx/dt and its acceleration is d^2x/dt^2 When the point with coordinates (x, y) moves in the xy

plane its velocity and acceleration are vectors the x and y components of velocity are dx/dt and dy/dt respectively and the corresponding components of acceleration are d^2x/dt^2 d^2y/dt^2 These results are extended naturally for a point $(x\ y\ z)$ moving in three dimensional space

For the plane case if the point has polar coordinates $(r\ \theta)$ the velocity and acceleration can be resolved into components in the r and θ directions The r direction at $(r\ \theta)$ is the direction directly away from the origin (r increasing and θ constant) The θ direction is 90° from the r direction in the counterclockwise sense The r and θ components of velocity are respectively $dr\ dt$ and $r(d\theta/dt)$ assuming that $r > 0$ The corresponding components of acceleration are

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 \quad \text{and} \quad r\frac{d^2\theta}{dt^2} + 2\frac{dr}{dt}\frac{d\theta}{dt}$$

These expressions are useful in discussing the motion of a particle under the influence of a central force as in the case of a single mass moving in the gravitational field of a fixed center of attraction according to the inverse square law

Still another useful way of resolving acceleration into components involves the tangential and normal directions to the path Here one needs to deal with arc length and curvature (*see* PARAMETRIC EQUATION) If s is arc length measured along a plane curve in a preassigned positive sense and if K is curvature then the velocity is a vector whose component along the curve is ds/dt the component at right angles to the curve being zero The component of acceleration along the curve in the positive sense is d^2s/dt^2 and the component at right angles to the curve (90° counter clockwise from the positive sense along the curve) is v^2K where v ds/dt The curvature of a circle is the reciprocal of its radius R and so a point moving around a circle of radius R with speed v experiences an acceleration toward the center of amount v^2/R This is called centripetal acceleration If the speed is not uniform there is also an acceleration along the curve

Curve tracing If $f'(x) > 0$ $f(x)$ increases as x increases whereas $f(x)$ decreases as x increases when $f'(x) < 0$ Points where $f'(x) = 0$ are called critical points With x and y axes in the usual position (x positive to the right y positive toward the top of the page) the graph of $y = f(x)$ is called concave upward over an interval of x values if the tangent line turns counterclockwise as x increases If the tangent line turns clockwise as x increases the curve is called concave downward If $f''(x) > 0$ the curve is concave upward and it is concave downward if $f''(x) < 0$ It follows that y is at a relative maximum if $f'(x) = 0$ and $f''(x) < 0$ at a relative minimum if $f'(x) = 0$ and $f''(x) > 0$ A point where the concavity changes from one sense to the other is called a point of inflection Such a point is a point of relative maximum or minimum for $f'(x)$ A sufficient condition for a point of inflection is that $f''(x) = 0$ $f'''(x) \neq 0$

Simple harmonic motion A point moving on the x axis in such a way that

$$\frac{d^2x}{dt^2} = -\omega^2 x$$

where $\omega > 0$ is executing what is called simple harmonic motion If a point travels around a circle with constant speed its orthogonal projection on a diameter executes simple harmonic motion with the center of the circle as the origin of coordinates on the diameter The methods of calculus enable one to infer from the foregoing differential equation that x depends on t by a formula $x = A\cos(\omega t + \alpha)$ where A and α are constants The moving point completes one cycle of its motion in time $2\pi/\omega$

Law of the mean For a discussion of the simple case of this law *see* CALCULUS DIFFERENTIAL AND INTEGRAL An extended version often called Cauchy s formula is the following Suppose F and G are continuous when $a \leqq x \leq b$ differentiable when $a < x < b$ and suppose $G(b) \neq G(a)$ Suppose also that $F'(x)$ and $G'(x)$ are never both zero together Then there is some number X such that $a < X < b$ and

$$\frac{F(b) - F(a)}{G(b) - G(a)} = \frac{F'(X)}{G'(X)}$$

Rule of l Hospital This rule for finding the limit of a quotient of two functions in certain circumstances is stated as follows Subject to certain general conditions on f and g if either

$$f(x) \to 0 \quad \text{and} \quad g(x) \to 0 \quad \text{as} \quad x \to a$$

or $g(x) \to +\infty$ or $g(x) \to -\infty$ as $x \to a$

then $$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

provided the limit on the right exists as a finite limit or as either $+\infty$ or $-\infty$ The general conditions are that f and g are differentiable as $x \to a$ and neither $g(x)$ nor $g'(x)$ is 0 as $x \to a$

Taylor s formula This formula is a finite sum expression for $f(x)$

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n + R_n$$

where R_n is called the remainder The two most useful formulas for R_n are

$$R_n = \frac{1}{n!}\int_a^x f^{(n+1)}(t)(x - t)^n\,dt \quad \text{(integral form)}$$

and

$$R_n = \frac{f^{(n+1)}(X)}{(n+1)!}(x - a)^{n+1} \quad \text{(Lagrange s f}$$

...atter form X is some number ... for fixed a and x $R_n \to 0$

$f(x)$ is said to be represented by Taylor s infinite series expansion (*see* SERIES) *See also* INTEGRATION, OPERATOR THEORY [A E T]

Diffraction

The bending of light or other waves into the region of the geometrical shadow of an obstacle More exactly diffraction refers to any redistribution in space of the intensity of waves that results from the presence of an object that causes variations of either the amplitude or phase of the waves Most diffraction gratings cause a periodic modulation of the phase across the wavefront rather than a modulation of the amplitude Although diffraction is an effect exhibited by all types of wave motion this article will deal only with electromagnetic waves especially those of visible light Some important differences that occur with microwaves will also be mentioned For discussion of the phenomenon as encountered in other types of waves *see* ELECTRON DIFFRACTION NEUTRON DIFFRACTION SOUND

Diffraction is a phenomenon of all electromagnetic radiation including radio waves microwaves infrared visible and ultraviolet light and x rays In the last case it shows special features that are appropriately discussed elsewhere (*see* X RAY DIFFRACTION) For a discussion of diffraction of radio waves *see* RADIO WAVE PROPAGATION The effects for light are important in connection with the resolving power of optical instruments

Classes There are two main classes of diffraction known as Fraunhofer diffraction and Fresnel diffraction The former concerns beams of parallel light and is distinguished by the simplicity of the mathematical treatment required and also by its practical importance The latter class includes the effects in divergent light and is the simplest to observe experimentally A complete explanation of Fresnel diffraction has challenged the most able physicists although a satisfactory approximate account of its main features was given by A Fresnel in 1814 At that time it played an important part in establishing the wave theory of light

To illustrate the difference between the methods of observation of the two types of diffraction Fig 1 shows the experimental arrangements required to ... screen S The

In Fraunhofer diffraction, the source lies at the principal focus of a lens L_1 which renders the light parallel as it falls on the aperture A second lens L_2 focuses parallel diffracted beams on the observing screen F situated in the principal focal plane of L_2 In Fresnel diffraction no lenses intervene The diffraction effects occur chiefly near the borders of the geometrical shadow, indicated by the broken lines An alternative way of distinguishing the two classes therefore is to say that Fraunhofer diffraction concerns the effects near the focal point of a lens or mirror while Fresnel diffraction concerns those effects near the edges of shadows Photographs of some diffraction patterns of each class are shown in Fig 2

FRAUNHOFER DIFFRACTION

This class of diffraction is characterized by a linear variation of the phases of the Huygens secondary waves with distance across the wavefront as they arrive at a given point on the observing screen (*see* HUYGENS' PRINCIPLE) At the instant that the incident plane wave occupies the plane of the diffracting screen it may be regarded as sending out from each element of its surface a multitude of secondary waves the joint effect of which is to be evaluated in the focal plane of the lens L_2 The analysis of these secondary waves involves taking account of both their amplitudes and their phases The simplest way to do this is to use a graphical method the method of the so-called vibration curve which can readily be extended to cases of Fresnel diffraction

Vibration curve The basis of the graphical method is the representation of the amplitude and phase of a wave arriving at any point by a vector the length of which gives the magnitude of the

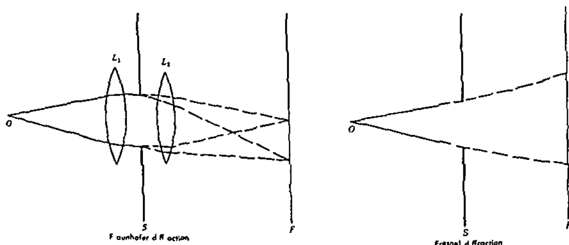


Fig 1 Observation of the two principal types of diffraction in the case of a circular aperture

Fig 2 Diffraction patterns, photographed with visible light (a) Fraunhofer pattern, for a slit (b) Fraunhofer pattern, square aperture, short exposure, (c) Fraunhofer pattern square aperture, long exposure (F S Harris) (d) Fraunhofer pattern, circular aperture (R W Ditchburn), (e) Fresnel pattern, straight edge (f) Fresnel pattern circular aperture

amplitude, and the slope of which gives the value of the phase In Fig 3 are shown two vectors of amplitudes a_1 and a_2, pertaining to two waves having a phase difference δ of 60° That is the waves differ in phase by one sixth of a complete vibration The resultant amplitude A and phase θ (relative to the phase of the first wave) are then found from the vector sum of a_1 and a_2 as indicated A mathematical proof shows that this proposition is rigorously correct and that it may be extended to cover the addition of any number of waves

The vibration curve results from the addition of a large (really infinite) number of infinitesimal vectors, each representing the contribution of the Huygens secondary waves from an element of surface of the wave front If these elements are assumed to be of equal area, the magnitudes of the amplitudes to be added will all be equal They will however, generally differ in phase so that if the elements were small but finite each would be drawn at a small angle with the preceding one as shown in Fig 4*a* The resultant of all elements would be the vector A When the individual vectors represent the contributions from infinitesimal surface elements (as they must for the Huygens wavelets), the diagram becomes a smooth curve the vibration curve, shown in Fig 4*b* The intensity on the screen is then proportional to the square of this resultant amplitude In this way, the distribution of the intensity of light in any Fraunhofer diffraction pattern may be determined

The vibration curve for Fraunhofer diffraction by screens having slits with parallel, straight edges is a circle Consider, for example the case of a slit of width b illustrated in Fig 5 The edges of the slit extend perpendicular to the plane of the figure and the slit is illuminated by plane waves of light coming from the left If s is the distance from O of a surface element ds (ds actually being a strip extending perpendicular to the figure) the extra distance that the wavelet from ds must travel in reaching a point on the screen lying at the angle θ

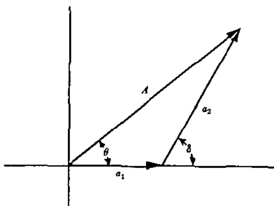


Fig 3 Graphical addition of two amp

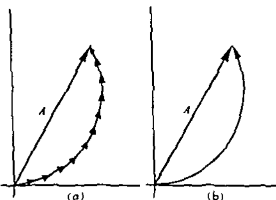


Fig 4 (a) Addition of many equal amplitudes differing in phase by equal amounts (b) Equivalent vibration curve when amplitudes and phase differences become infinitesimal

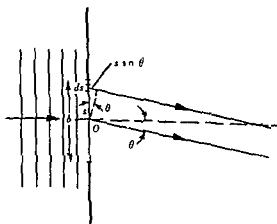


Fig 5 Analysis of Fraunhofer diffraction by a slit

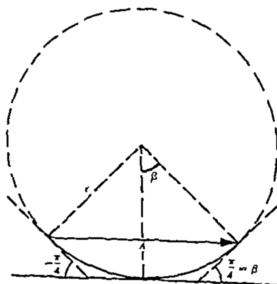


Fig 6 Vibration curve and resultant amplitude for a particular point in Fraunhofer diffraction by a slit

from the center is $s \sin \theta$ Since this extra distance determines the phase difference, the latter varies linearly with s This condition necessitates that the vibration curve be a circle

The intensity distribution for Fraunhofer diffraction by a slit as a function of the angle θ may be simply calculated as follows The extra distances traveled by the wavelets from the upper and lower edges of the slit, as compared with those from the center are $+(b/2) \sin \theta$ and $-(b/2) \sin \theta$ The corresponding phase differences are $2\pi/\lambda$ times these quantities, λ being the wavelength of the light Using the symbol β for $(\pi b \sin \theta)/\lambda$ it is seen that the end points of the effective part of the vibration curve must differ in slope by $\pm\beta$ from the slope at its center, where it is taken as zero Figure 6 shows the form of the vibration curve for $\beta = \pi/4$ that is, for $\sin \theta = \frac{1}{4}\lambda/b$ The resultant is $A = 2r \sin \beta$ where r is the radius of the arc The amplitude A_0 that would be obtained if all the secondary waves were in phase at the center of the diffraction pattern where $\theta = 0$ is the length of the arc Thus,

$$\frac{A}{A_0} = \frac{\text{chord}}{\text{arc}} = \frac{2r \sin \beta}{2r\beta} = \frac{\sin \beta}{\beta}$$

The intensity at any angle is

$$I = I_0 \frac{\sin^2\beta}{\beta^2} \tag{1}$$

where I_0 is the intensity at the center of the pattern Figure 7 shows a graph of this function The central maximum is twice as wide as the subsidiary ones and is about 21 times as intense as the strongest of these A photograph of this pattern is shown in Fig 2a

The dimensions of the pattern are important since they determine the angular spread of the light behind the slit The first zeros occur at values $\beta = (\pi b \sin \theta)/\lambda = \pm\pi$ In most cases the angle θ is extremely small so that

$$\sin \theta = \theta = \pm\frac{\lambda}{b} \tag{2}$$

For a slit 1 mm wide, for example, and green light of wavelength 5×10^{-5} cm Eq (2) gives the angle as only 0 0005 radians, or 1 72 minutes of arc The slit would have to be much narrower than this or the wavelength much longer, for the approximation to cease to be valid

The main features of Fraunhofer diffraction patterns of other shapes can be understood with the aid of the vibration curve Thus for a rectangular or square aperture, the wavefront may be subdivided into elements parallel to either of two adjacent sides giving an intensity distribution which follows the curve of Fig 7 in the directions parallel to the two sides Photographs of such patterns appear in Fig 2b and c In Fig 2c it will be seen that there are also faint subsidiary maxima lying off the two principal directions These have

intensities proportional to the products of the intensities of the side maxima in the slit pattern The fact that these subsidiary intensities are extremely low compared with that of the central maximum has an important application to the apodization of lenses, to be discussed later

Diffraction grating. An idealized diffraction grating consists of a large number of similar slits equally spaced Equal segments of the vibration curve are therefore effective, as shown in Fig 8*a* The resultants *a* of each segment are then to be added to give A, the amplitude due to the whole grating as shown in Fig 8*b* The phase difference between successive elements is here assumed to be very small As it is increased by going to a larger angle θ, the resultant A first goes to zero at an angle corresponding to λ/W, where W is the total width of the grating After going through numerous low intensity maxima A again rises to a high value when the phase difference between the successive vectors for the individual slits approaches a whole vibration These small vectors a are then all lined up again as they were at the center of the pattern ($\theta = 0$) The resulting strong maximum represents the "first order spectrum" since its position depends on the wavelength A similar condition occurs when the phase difference becomes two three or more whole vibrations giving the higher order spectra By means of this diagram it is possible not only to predict the intensities of successive orders for an ideal grating but also to find the sharpness of the maxima which represent the spectrum lines *See* DIFFRACTION GRATING

Determination of resolving power Fraunhofer diffraction by a circular aperture determines the resolving power of instruments such as telescopes

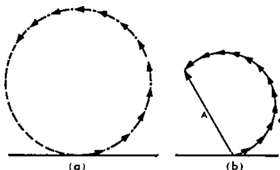


Fig 8 (a) Vibration curve for diffraction grating of twelve slits (b) Resultant amplitude A formed by adding amplitudes a from individual slits Each a represents the chord of one of the short arcs in part (a)

cameras and microscopes in which the width of the light beam is usually limited by the rim of one of the lenses The method of the vibration curve may be extended to find the angular width of the central diffraction maximum for this case Figure 9 compares the treatments of square and circular apertures by showing above the elements of equal phase difference into which the wavefront may be divided and below the corresponding vibration curves For the square aperture shown in Fig 9*a* the areas of the surface elements are equal and the curve forms a complete circle at the first zero of intensity In Fig 9*b* these areas and hence the lengths of the successive vectors are not equal but increase as the center of the curve is approached and then decrease again The result is that the curve must show somewhat greater phase differences at its extremes in order to form a closed fig

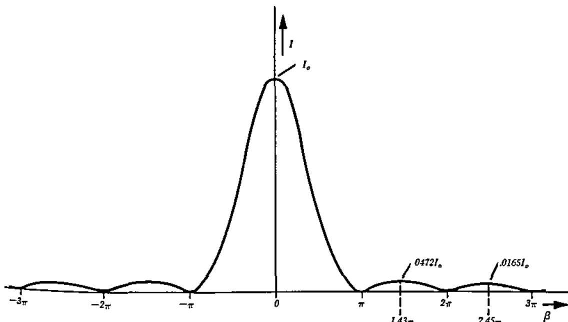
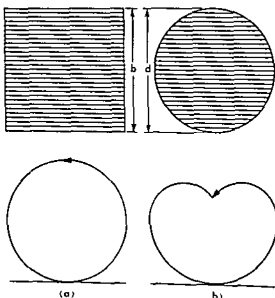


Fig 7 Intensity distribution for Fraunhofer diffraction by a slit



F g 9 Cond t on of the v brat on curves at the first min mum for (a) a square aperture and (b) a circular aperture

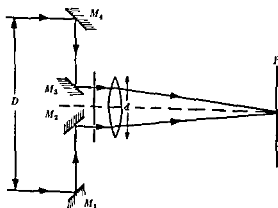


Fig 10 Michelson stellar interferometer used for measur ng d ameters of stars

ure An exact . .

e angle θ at the first zero of intensity

$$\sin\theta \approx \theta = \pm\frac{1.220\lambda}{d} \qquad (3)$$

Here d is the diameter of the circular aperture When this result is compared with that of Eq (2) it is apparent that relative to the pattern . . .

. . . called the Airy disk surrounded by faint rings (see Fig 2d) The angular radius of the disk given by Eq (3) may be extremely small for an actual optical instrument but it sets the ultimate limit to the sharpness of the image that is to the resolving power See RESOLVING POWER (OPTICS)

Other applications Among the applications of Fraunhofer diffraction, other than the calculation of resolving power are its use in (1) the theory of certain types of interferometer in which the interfering light beams are brought together by diffraction (2) the theory of microscopic imaging and (3) the design of apodizing screens for lenses

Michelson's stellar interferometer This instrument was devised by A A Michelson to overcome the limitation expressed by Eq (3) In front of a telescope are placed two fixed mirrors shown at M_2 and M_3 in Fig 10 and two others M_1 and M_4 movable so as to vary their separation D The light of a star is directed by these mirrors into a telescope as shown and the diffraction patterns of size corresponding to the aperture of the mirrors are superimposed in the focal plane P of the lens When D is small the resulting pattern is crossed by interference fringes If the light source has a finite width increasing D causes these fringes to become indistinct and eventually to disappear For a circular disk source such as a star the angle subtended by the disk must be $1.22\lambda/D$ for disappearance of the fringes The resolving power of the telescope has thus effectively been increased in the ratio D/d In this way angular diameters as small as 0.02 seconds of arc have been measured corresponding to D = 24 ft *See* INTERFERENCE OF WAVES, INTERFEROMETRY

Microscopic imaging F Abbe's theory of the microscope evaluates the resolving power of this instrument by considering not only the diffraction caused by the limited aperture of the objective but also that caused by the object itself If this object is illuminated by coherent light such as a parallel beam from a point source its Fraunhofer diffraction pattern is formed in the rear focal plane of the microscope objective Abbe took as an object a diffraction grating and in this case the pattern consists of a series of sharp maxima representing the various orders $m = 0, \pm1, \pm2$ If the light of all these orders is collected by the objective a perfect image of the grating can be formed where the light is reunited in the image plane In practice however the objective can include only a limited number of them as is indicated in Fig 11 Here only the orders +2 to −2 are shown entering the objective The higher orders involving greater values of θ would miss the lens

The final image must be produced by interference in the image plane of the Huygens secondary waves coming from the various orders To obtain a periodic variation of intensity in that plane at least two orders must be accepted by the objective The angle θ_1 at which the first order occurs is given by $\sin\theta_1 = \lambda/d$ where d is the spacing of the lines in the grating The limit of resolution of the microscope that is the smallest value of d that will produce an indication of separated lines in the image may thus be found from the angular aperture of the objective In order for it to accept only the orders 0 ±1 the lens aperture 2α must equal at least $2\theta_1$ giving $d = \lambda/\sin\alpha$ This resolving

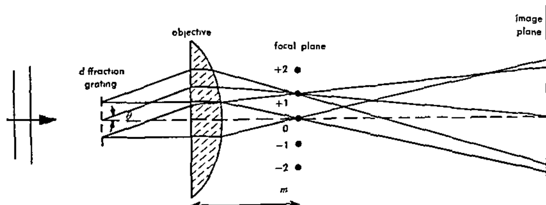


Fig 11 Abbe's method of treating the resolving power of a microscope

limit may be decreased by illuminating the grating from one side so that the zero order light falls at one edge of the lens and that of one of the first orders at the other edge Then $\theta_1 = 2\alpha$ and the limit of resolution is approximately 0 5λ sin α *See* MICROSCOPE OPTICAL

Apodization This is the name given to a procedure by which the effect of subsidiary maxima (such as those shown in Fig 7) may be partially suppressed Such a suppression is desirable when one wishes to observe the image of a very faint object adjacent to a strong one If the fainter object has for example only 1/1000 of the intensity of the stronger one the two images will have to be far enough apart so that the principal maximum for the fainter one is at least comparable in intensity to the secondary maxima for the stronger object at that point In the pattern of a rectangular aperture it is not until the tenth secondary maximum that the intensity of these falls below 1 1000 of the intensity of the principal maximum Here it has been assumed however that the fainter image lies along one of the two principal directions of diffraction of the rectangular aperture perpendicular to two of its adjacent sides At 45° to these directions the subsidiary maxima are much fainter (see Fig 2c) and even the second one has an intensity of only 1/3700 the square of the value of the second subsidiary maximum indicated in Fig 7

The simplest apodizing screen is a square aperture placed over a lens the diagonal of the square being equal to the diameter of the lens If the lens is the objective of an astronomical telescope for example the presence of a fainter companion in a double star system can often be detected by turning the square aperture until the image of the companion star lies along its diagonal Apodizing screens may be of various shapes depending on the purpose to be achieved It has been found that a screen of graded density which shades the lens from complete opacity at the rim to complete transparency at a small distance inward is effective in suppressing the circular diffraction rings surrounding the Airy disk In all types of apodization there is some sacrifice of true resolving power so that it would not be used if the two images to be resolved were of equal intensity

FRESNEL DIFFRACTION

The diffraction effects obtained when the source of light or the observing screen are at a finite distance from the diffracting aperture or obstacle come under the classification of Fresnel diffraction This type of diffraction requires for its observation only a point source a diffracting screen of some sort and an observing screen The latter is often advantageously replaced by a magnifier or a low power microscope The observed diffraction patterns generally differ according to the radius of curvature of the wave and the distance of the point of observation behind the screen If the diffracting screen has circular symmetry such as that of an opaque disk or a round hole a point source of light must be used If it has straight parallel edges it is desirable from the standpoint of brightness to use an illuminated slit parallel to these edges In the latter case it is possible to regard the wave emanating from the slit as a cylindrical one For the purpose of deriving the vibration curve the appropriate way of dividing the wavefront into infinitesimal elements is to use annular rings in the first case and strips parallel to the axis of the cylinder in the second case

Figure 12 illustrates the way in which the radii of the rings or the distances to the edges of the strips must be chosen in order that the phase difference may increase by an equal amount from one element to the next Figure 12*a* shows a section of the wavefront diverging from the source S and the paths to the screen of two secondary wavelets The shortest possible path is b while r is that for another wavelet originating at a distance s above the pole O Since all points on the wavefront are at the same distance a from S the path difference between the two routes from S to P is $r - b$ When this is evaluated to terms of the first order in s/a and s/b the phase difference is

$$\delta = \frac{2\pi}{\lambda}(r-b) = \frac{\pi(a+b)}{ab\lambda}s^2 = Cs^2 \qquad (4)$$

The phase difference across an elementary zone of radius s and width ds then becomes $d\delta = 2Cs\,ds$ so that for equal increments of δ the increment of s must be proportional to $1/s$ The annular zones and

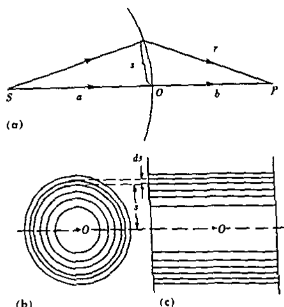


Fig 12 Division of wavefront for constructing vibration curve for Fresnel diffraction (a) Section of wave diverging from S and paths to screen of two secondary wavelets (b) Annular zones (c) Strips

the strips drawn in this way on the spherical or cylindrical wave respectively looking toward the pole from the direction of P are shown in Fig 12b and Fig 12c

For the annular zonal elements the areas $2\pi s ds$ are all equal and hence the amplitude elements of the vibration curve should have the same magnitude Actually they must be regarded as falling off slowly due to the influence of the "obliquity factor" of Huygens' principle The resulting vibration curve is nearly but not quite circular and is illustrated in Fig 13a It spirals in toward the center C at a rate that has been considerably exaggerated in the figure The intensity at any point P on the axis of a circular screen centered on O can now be determined as the square of the resultant amplitude A for the appropriate part or parts of this curve The curve shown in Fig 13a is for a circular aperture which exposes five Fresnel zones each one represented by a half turn of the spiral The resultant amplitude is almost twice as great (and the intensity four times as great) as it would be if the whole wave were exposed in which case the vector would terminate at C The diffraction by other circular screens may be determined in this way

Zone plate This is a special screen designed to block off the light from every other half period zone and represents an interesting application of Fresnel diffraction The Fresnel half period zones are drawn with radii proportional to the square roots of whole numbers and alternate ones are blackened The drawing is then photographed on a reduced scale When light from a point source is sent through the negative an intense point image is produced much like that formed by a lens The zone plate has the effect of removing alternate half turns of the spiral, the resultants of the others all adding in the same phase By putting $\delta = -$ and $a = \infty$ in Eq (4) it is found that the "focal length" b of a zone plate is s_1^2/λ, where s_1 is the

for a
13b

The areas of the elementary zones, and hence the magnitudes of the component vectors of the vibration curve decrease rapidly, being proportional to ds, and hence to $1/s$ (see Fig 12c) The definition of Cornu's spiral requires that its slope δ at any point be proportional to the square of the corresponding distance s measured up the wavefront [see Fig 12a and Eq (4)] The length of the spiral from the origin is proportional to s, but it is usually drawn in terms of the dimensionless variable v defined by $v = s\sqrt{2C/\pi}$ in the notation of Eq (4) The coordinates of any point on the curve may then be found from tables of Fresnel's integrals

As an example of the application of Cornu's spiral consider the diffraction of an opaque straight edge such as a razor edge illuminated by light from a narrow slit parallel to the edge At some point outside the edge of the shadow, say one which exposes three half period strips beyond the pole the resultant amplitude will be that labeled A in Fig 13b The intensity will be greater than that given by the amplitude NN', which represents the

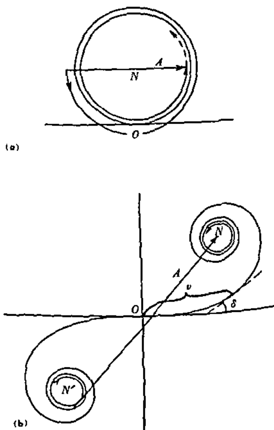


Fig 13 Vibration curves for Fresnel diffraction (a) Circular division of wavefront (b) Strip division (Cornu's spiral)

amplitude for the whole (unobstructed) wave On going further away from the edge of the shadow the tail of the vector A will move along the spiral inward toward N and the intensity will pass through maxima and minima At the edge of the geometrical shadow the amplitude is ON and the intensity is just one fourth of that due to the un obstructed wave Further into the shadow the in tensity approaches zero regularly without fluctua tions as the tail of the vector moves up toward N A photograph of the straight edge pattern is shown in Fig 2e

Babinet s principle This states that the diffrac tion patterns produced by complementary screens are identical Two screens are said to be comple mentary when the opaque parts of one correspond to the transparent parts of the other and vice versa Babinet s principle is not very useful in dealing with Fresnel diffraction except that it may furnish a short cut method in obtaining the pattern for a particular screen from that of its complement The principle has an important application for Fraun hofer diffraction in parts of the field where there is zero intensity without any screen Under this con dition the amplitudes produced by the complemen tary screens must be equal and opposite since the sum of the effects of their exposed parts gives no light The intensities being proportional to the squares of the amplitudes must therefore be equal In Fraunhofer diffraction the pattern due to a disk is the same as that due to a circular hole of the same size

DIFFRACTION OF MICROWAVES

The diffraction of microwaves which have wave lengths in the range of millimeters to centimeters has been intensively studied since World War II because of its importance in radar work Many of the characteristics of optical diffraction can be strikingly demonstrated by the use of microwaves Microwave diffraction shows certain features how ever that are not in agreement with the Huygens Fresnel theory because the approximations made in that theory are no longer valid Most of these approximations as for example that made in de riving Eq (4) depend for their validity on the assumption that the wavelength is small compared to the dimensions of the apparatus Furthermore it is not legitimate to postulate that the wave has a constant amplitude across an opening and zero intensity behind the opaque parts except for the very minute waves of light

As an example of the failure of classical diffrac tion theory when applied to microwaves the results of the diffraction by a circular hole in a metal screen may be mentioned The observed patterns begin to show deviations from the Fresnel theory and even from the more rigorous Kirchhoff theory when the point of observation is within a few wave lengths distance from the plane of the aperture Even in this plane itself there are detectable varia tions of intensity There are also polar ef fects that could not have been predicted earlier theories which treat light as a rather than a vector wave motion An exact vector theory of diffraction developed by A Sommerfeld has been applied only in a few simple cases but the measurements at microwave frequencies agree with it wherever it has been tested See MICROWAVE OPTICS see also LIGHT [F A J]

Bibliography F A Jenkins and H E White *Fundamentals of Optics* 3d ed 1957 C F Meyer *The Diffraction of Light X rays and Material Particles* 1934 A Sommerfeld *Lectures on Theoretical Physics* vol 5 1954

Diffraction grating

An optical device consisting of an assembly of nar row slits or grooves which by diffracting light produces a large number of beams which can inter fere in such a way as to produce spectra Since the angles at which constructive interference patterns are produced by a grating depend on the lengths of the waves being diffracted the waves of various lengths in a beam of light striking the grating will be separated into a number of spectra produced in various orders of interference on either side of an undiffracted central image By controlling the shape and size of the diffracting grooves when producing a grating and by illuminating the grat ing at suitable angles a beam of light can be thrown into a single spectrum whose purity and brightness may exceed that produced by a prism Gratings can now be made with much larger aper tures than prisms and in such form that they waste less light and give higher intrinsic dispersion and resolving power A single grating can be used over a much broader range of spectrum than can any single prism and its dispersion will vary less rap idly with wavelength Gratings are being used in creasingly in large spectrographs and for highly precise spectroscopic work as well as in mono chromators and analytical spectrographs See DIFFRACTION INTERFERENCE OF WAVES PRISM OPTICAL SPECTROSCOPY

Transmission gratings consist of a large number of narrow transparent and opaque slits alternating side by side in regular order and with uniform separation through which a beam of light will appear as a series of spectra in various orders of interference Such gratings are conveniently used in small spectroscopes and spectrometers but only for visible light since they are usually not trans parent to ultraviolet or infrared radiation They are commonly made by contact molding from a master grating

Reflection gratings either plane or concave are used in most spectrographs Such a grating may consist of an original ruling or of a metal coated replica from an original Large grating replicas can now be made which are practically indistin guishable in performance or permanence from an original

Production of gratings Gratings are engraved by highly precise ruling engines which use a dia mond tool to press into a highly polished mirror surface a series of many thousands of fine burnished grooves Gratings for the range

10 000 A are commonly ruled with 5000–30 000 grooves per inch (the usual value is near 15 000) on a thin layer of aluminum deposited on glass by evaporation in vacuum Gratings for the infrared region are often ruled on gold silver copper lead or tin mirrors with coarser groove spacings

If a grating is to give resolution approaching the theoretical limit its grooves must be ruled straight parallel and equally spaced to within a few tenths of the shortest incident wavelength The proper over all spacing of grooves must also be maintained if changes in focal properties are not to result Scattered light and false images may arise from local spacing error and groove shape variations of only a few hundredths of the diffracted wavelength

Among the false lines produced by imperfect gratings are Rowland ghosts which arise from periodic errors in groove position Lyman ghosts which come from a combination of periodicities in ruling and satellites caused by sets of irregularly placed grooves which may seriously reduce resolution Target pattern arises from unequal contribution of light from all parts of a grating and is especially prevalent in concave gratings in which the shape of the grooves may change as the cutting angle of the diamond changes

Gratings of 2 in 4-in or 6 in ruled width are commonly used in commercial spectrographs with projection distances of 20–180 in In large research instruments gratings of 6 to 10 in ruled width are used with projection distances of 10–50 ft or more The largest modern gratings used in their highest orders show resolving power $\lambda/\delta\lambda$ in excess of 900 000 in the green region of the spectrum and in excess of $1\,5 \times 10^6$ at shorter wavelengths Here λ is the mean wavelength of two closely spaced just resolvable spectral lines and $\delta\lambda$ is their wavelength difference Such gratings give resolution equal to that of most interferometers and in addition provide greater photographic speed are easier to adjust follow more simple laws of wavelength distribution and permit a wider range of wavelengths to be photographed at one time without crossed dispersion *See* INTERFEROMETRY RESOLVING POWER (OPTICS)

Properties of gratings A grating illuminated at angle α (measured from the normal) will direct wavelength λ toward angle β in accordance with the formula $m\lambda = d(\sin\alpha \pm \sin\beta)$ where m is an integral order of interference d is the grating constant or distance between consecutive grooves and the + and − signs refer to orders on opposite sides of the normal The linear dispersion produced by a grating on a photographic plate depends on its intrinsic angular dispersion multiplied by the distance P from grating to plate The intrinsic angular dispersion is given by the formula

$$\frac{d\beta}{d\lambda} = \frac{1}{\lambda}\left(\frac{\sin\alpha}{\cos\beta} + \tan\beta\right)$$

Theoretically the resolving power of a grating is mN where N is the number of grooves in the grating Resolving power is not directly dependent on the number of grooves since for gratings of a given size m and N are inversely related It is basically dependent on the number of wavelengths of optical retardation the grating introduces between the extreme rays leaving it Another useful concept is the resolving limit $d\sigma$ the smallest wavenumber difference the grating can resolve which remains essentially constant for a given angle of illumination of a given grating at all wavelengths except as errors in groove spacing become of greater importance for the shorter wavelengths

The manner in which incident light will be distributed among the various orders of interference depends upon the shape and orientation of the groove sides and on the relation of wavelength to groove separation When $d \gtrsim \lambda$ diffraction effects predominate in controlling the intensity distribution among orders but when $d > \lambda$ optical reflection from the sides of the grooves is more strongly involved It is possible to blaze a grating by ruling its grooves so that their sides reflect a large fraction of the incoming light of suitably short wavelengths in one general direction Controlled groove shape is especially important in the gratings known as echelettes and echelles in which as much as 80% of the incoming light may be sent into one particular order for a given wavelength Many ordinary gratings are blazed

Grating spectroscopes These consist usually of a slit a lens or mirror to collimate the light sent through the slit into a parallel beam a transmission or reflection grating to disperse the light a lens or mirror to focus the light into spectrum lines (which are monochromatic images of the slit in the light of each wavelength passing through it) and an eyepiece for viewing the spectrum (see Fig 1) If a camera is substituted for the telescope the instrument becomes a grating spectrograph If a photoelectric cell a thermocouple or other radiation detecting device is used instead of a camera or telescope the device becomes a grating spectrometer For some important applications of the latter device *see* INFRARED SPECTROSCOPY

Echelette grating This has coarse groove spacing and is designed for the infrared region the grooves being so shaped and of such size that most of the radiation is concentrated by reflection into a small angular coverage Radiation of any given wavelength is thus concentrated largely into one order by shaping the point of the ruling diamond to give

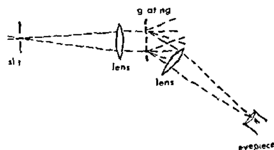


Fig 1 Transmission grating spectroscope

grooves with comparatively flat sides and choosing a groove separation which minimizes diffraction effects

Echelle grating This is designed for use in high orders and at angles of illumination greater than 45° to obtain high dispersion and resolving power by the use of high orders of interference Echelles have properties lying midway between those of plane gratings of the ordinary type and interferometers of the reflection echelon type orders of interference ranging from 100 to 1000 being used Overlapping orders are separated by using crossed dispersion

Concave grating This is a widely used form of reflection grating with which a spectrograph can be formed that has no auxiliary optical parts except a slit and a camera Being ruled on a concave mirror this type of grating can both collimate and focus the light that falls upon it It is made by spacing straight grooves equally along the chord (rather than the arc) of a spherical or paraboloidal mirror surface Light which passes through a slit and falls on such a grating is dispersed by it into spectra which are in focus on the Rowland circle a circle drawn tangent to the face of the grating at its midpoint having a diameter equal to the radius of curvature of the grating surface

A great advantage of the concave grating is that it provides a dispersing and focusing system free of refracting material so that it can be used with ultraviolet visible or infrared radiation interchangeably so long as its grooves diffract radiation and its surface has adequate reflecting power A disadvantage is its astigmatism at high angles of incidence or reflection which can however be diminished with various optical devices A plane reflection grating used with two concave mirrors avoids this difficulty

Grating mountings The slit grating and camera of a concave grating spectrograph can be placed anywhere on the Rowland circle so that any desired wavelength range can be photographed in the desired order (see Fig 2)

The various possible combinations of fixed and moving parts give rise to a number of different grating mountings In the Rowland mounting camera and grating are connected by a bar forming a diameter of the Rowland circle the two running on tracks placed at right angles with the slit fixed at their junction A spectrum of limited extent having uniform dispersion is then produced at the camera and camera and grating can be moved on the tracks to shift wavelength coverage In the Paschen Runge mounting slit and grating are fixed and photographic plates can be clamped to a fixed track almost anywhere on the Rowland circle In the Eagle mounting most suited to long narrow housing the grating can be rotated and moved toward or away from the slit The slit is placed close to the camera which is arranged to rotate so that it can be kept on the Rowland circle

All these mountings suffer from astigmatism arising from using the grating off axis Although this does not markedly reduce the sharpness of the spectrum lines it may result in a great loss of light intensity when the grating is used at high angles and it makes difficult the sharp focusing of step filters sector disks and interferometer patterns that are placed at the slit

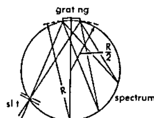


Fig 2 The Rowland circle

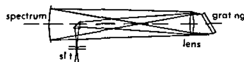


Fig 3 Littrow mounting of a plane grating

Astigmatism is greatly reduced in the Wadsworth mounting in which the slit is placed at the principal focus of a concave mirror so that the light falling on the grating is in a parallel beam The grating can be illuminated at any desired angle up to about 40° and light is taken off along the grating normal the spectrum being focused on a photographic plate at only half the usual distance The usual dispersion of the grating is thus cut in half but the speed of the spectrograph is increased fourfold and at high angles the speed is increased much more as a result of reduction of astigmatism

Most modern commercial grating spectrographs because of the need for portability are based either on the Eagle mounting with which a rather limited portion of the spectrum can be photographed at one time or the Wadsworth mounting which can give greater spectral coverage without resetting but is bulkier and cannot be used at such high values of $m\lambda$ To obtain more complete spectrum coverage in a single exposure an echelle with crossed dispersion or a grating with some other device for separating orders may be used

Plane reflection gratings are ordinarily used in the Littrow mounting (Fig 3) in which a single lens serves for both collimating and focusing or in the Ebert mounting with two concave mirrors

[C R H]

Bibliography W R Brode *Chemical Spectroscopy* 2d ed 1943 G R Harrison R C Lord and J R Loofbourow *Practical Spectroscopy* 1948 R A Sawyer *Experimental Spectroscopy* 2d ed 1951

Diffuser

A device to convert a high velocity low pressure stream of fluid (usually air) into a low velocity high pressure flow In doing so a diffuser converts the velocity or kinetic energy into pressure or potential energy

Efficiency An efficient diffuser operates by decelerating the high velocity stream with as little loss in total energy as possible Although the efficiency of a diffuser may be described in many ways a convenient experimental measure of merit is the total pressure recovery that is the ratio of total pressure after diffusion to the total pressure in the stream ahead of the diffuser Total pressure is composed of ambient pressure and velocity head components (*see* PITOT TUBE) Another measure of performance is the kinetic energy efficiency which is the ratio of kinetic energy after diffusion to that before diffusion The kinetic energy (velocity energy) after diffusion is calculated by assuming isentropic reexpansion to free stream static (or ambient) pressure at diffuser exit total pressure Mathematically the two are related by

$$\eta_{KE} = 1 + \frac{2}{(\gamma - 1)M_0^2}\left[1 - \left(\frac{H_0}{H_1}\right)^{(\gamma-1)/\gamma}\right]$$

where η_{KE} is kinetic energy efficiency γ is ratio of specific heats M_0 is Mach number ratio of local stream velocity to velocity of sound and (H_1/H_0) is total pressure recovery

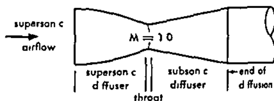
One dimensional representations of the flow are usually accurate enough to permit realistic or engineering evaluations of the experimental results However they will generally not rigorously satisfy the equations of conservation of mass momentum and energy simultaneously This is a consequence of the various distortions in velocity profile that can occur across the diffuser at any station thus making the one dimensional flow assumption less valid

Shape The shape of a diffuser depends on the Mach number of the entering fluid If a supersonic stream is to be decelerated the diffuser cross-sectional area must be progressively decreased (ideally to just sonic velocity) and then increased until the flow reaches the desired velocity as indicated in the sketch (*see* SUPERSONIC DIFFUSER) The progressive area change required of the flow is $dA/A = (du/u)(M - 1)$ where dA is the incre [illegible]

Mach number is supersonic (greater than 1.0) an area reduction is required to reduce the stream velocity the converse is true for an initial subsonic Mach number

As yet no restrictions have been placed by this one-dimensional representation on the rate at wh[illegible]

S[illegible] operation invariably results in low over all performance The cross sectional shape of the diffuser and the type of boundary layer further influence the permissible diffusion rate Optimum divergence angles are about 6° for conical or square



Idealized diffuser for decreasing supersonic velocities to subsonic velocities

diffusers and about 11° for two dimensional divergent channels

Applications Probably the most dramatic use of diffusers is on jet aircraft Diffusers supplying air to the engine have had inlets in the nose at the wing root along the airplane fuselage and even in the tail section In all these applications the same design guides apply for subsonic diffusers—reasonable diffuser angles maximum radius bends absence of protuberances in the duct and generously rounded inlet lips This last criterion conflicts with the requirements for supersonic inlets where sharp lips are desirable Consequently severe penalties may be incurred subsonically unless appropriate design compromises are effected

[J F CO , L J O]

Bibliography J C Evvard Diffusers and Nozzles in J V Charyk and M Summerfield (eds) *High Speed Aerodynamics and Jet Propulsion*, vol 7 1957

Diffusion in gases and liquids

The spreading or scattering of matter under the influence of a concentration gradient For a discussion of diffusion in solids *see* DIFFUSION IN SOLIDS

Molecular diffusion Consider a fluid confined in a space of dimensions which are large compared to the mean free path of the fluid molecules For gases at atmospheric pressure and room temperature the mean free path is on the order of 10^{-5} cm and varies inversely with pressure Liquids in general have much smaller free paths Assume that one of the components of the fluid exists initially at different concentrations or more rigorously at different activities in two or more different locations in the confining space At constant temperature and in the absence of external forces there will be a spontaneous movement that is diffusion of the component in the direction of establishing a uniform concentration of that component in all parts of the enclosure

In a very qualitative way the cause of this spontaneous mixing may be interpreted as follows As a consequence of thermal agitation molecules of a fluid are in constant motion in all directions The number of molecules of a given kind moving in any given direction at a particular point in the fluid is proportional to the number of these molecules present per unit volume In the absence of a concentration gradient on the average as many molecules per unit time leave any hypothetical plane in a given direction as return to the plane from the opposite direction However if the number of mole

cules per unit volume decreases in a given direction more molecules on the average move into the region of lower concentration than return from it This results in a net transfer of molecules of that particular kind toward the region of lower concentration

In general the diffusion of one component of a mixture will be accompanied by diffusion of one or more other components in the opposite direction to maintain the volume of fluid constant in the region under consideration This may not be the case when special restrictions are placed on the system for example in the selective absorption of one component from a gas by a liquid acting as a boundary of the confining space or for liquids when there is a volume change on mixing

The rate law for equivolume diffusion was proposed by Adolf Fick in 1855 For diffusion of component A in a mixture of A in component B in the z direction across a plane of unit cross section perpendicular to the direction of diffusion the rate is given by

$$N_{Ad} = -D_{AB}\frac{dc_A}{d} \qquad (1)$$

where N_{Ad} is the rate of diffusion in moles/(cm^2) (sec) D_{AB} is the diffusion coefficient of A in B cm^2/sec c_A is the concentration of component A moles/cm^3 and z is the distance in direction of diffusion in cm Similar equations may be written for concentration gradients in the other coordinate directions

At room temperature and atmospheric pressure diffusion coefficients for most gases and vapors lie in the range 0 1–1 cm^2/sec In liquids of about 1

turbulent Viscous flow is characterized by movement of the fluid in essentially smooth streamlines parallel to the walls of the pipe through which the fluid is flowing or more generally parallel to any surface over which flow occurs Turbulent flow on the other hand is characterized by movement of fluid in the form of statistically defined lumps or eddies which fluctuate randomly in velocity perpendicular to the surface as well as parallel to the direction of flow Turbulence develops when fluid flows past a surface at a sufficiently high velocity so that the Reynolds number exceeds some critical value for the system (*see* FLUID FLOW PRINCIPLES) In viscous flow mass transfer resulting from a concentration gradient occurs by molecular diffusion In turbulent flow transfer of material occurs by the more rapid process of mixing of the swirling eddies of fluid a process termed turbulent diffusion or eddy diffusion It is customary to define the mass transfer flux in eddy diffusion by means of an eddy diffusion coefficient E as follows

$$N_A = -E_A\frac{dc_A}{dz} \qquad (2)$$

where E is roughly proportional to the size of the eddies and to the magnitude of the velocity fluctuations Eddy diffusion coefficients are usually much larger than molecular coefficients In flow of air in a duct for example E has been observed to vary from 3 to 40 cm^2/sec over a range of Reynolds numbers from 10 000 to 175 000 Liquid phase values would be lower by a factor of about 10^{-3} However in the vicinity of a rigid boundary turbulence is suppressed and much lower values of E may exist so that the molecular and turbulent coefficients may become similar in magnitude and immediately at the interface E may become negligible Finally description of turbulent mixing with a simple diffusion law is a convenient oversimplification of an extremely complex fluid mechanical problem which is not fully understood

Thermal diffusion S Chapman demonstrated in 1916 as a consequence of the kinetic theory that a temperature gradient in a mixed gas might give rise to a flow of one constituent relative to the mixture as a whole His finding was verified experimentally by L Chapman and F W Dootson who placed various mixtures in a bulb which was heated at one end and cooled at the other A similar effect had been observed for liquids by Ludwig in 1856 and by J L Soret in 1879 and is usually called the Soret effect The phenomenon as it pertains to fluids generally is called thermal diffusion The thermal diffusion flux of a given component of a binary mixture is expressed by the equation

$$N_{AT} = D_{T}\rho\,\frac{d\ln T}{d} \qquad (3)$$

where N_A is the number of moles/(cm^2) (sec) D_T is the coefficient of thermal diffusion in cm^2/sec T is temperature in °K z is the distance in direction of diffusion in cm and ρ is the fluid density in g moles/cm^3

Magnitude of the thermal diffusion coefficient varies widely depending upon the sizes and chemical nature of the molecules However it seldom has a value greater than 30% of the molecular diffusion coefficient and is usually much smaller Hence unless temperature gradients are quite large and the fluid is nonturbulent as well thermal diffusion is not an important factor in most mass transfer operations Specific equipment and processes have been developed to utilize thermal diffusion for separation of isotopes and other substances that may not yield to less costly separation techniques *See* ISOTOPE SEPARATION (STABLE ISOTOPES)

Forced diffusion An external force acting upon each of a group of molecules may induce molecular movement analogous to diffusion Such movements which may result from gradients of pressure within a fluid or from electric or magnetic fields may be termed forced diffusion An important example is the process of electrolytic migration of ions under an applied potential Movement of materials between a working electrode and solution with which it is in contact may result predominantly from migration in absence of inert electrolyte to carry the

current. Rate of migration at a given point within a fluid is given by the equation

$$N_{A_m} = C_A U_A \frac{di}{dz} \quad (4)$$

where U_A is the mobility of A in cm²/(sec)(volt) and i is the potential in volts. Ionic mobilities depend upon the particular ion, the nature of the solvent, and the temperature. For dilute aqueous solutions at 25°C most ions with the exception of hydrogen and hydroxyl ions have mobilities in the range of 3–8 × 10⁻⁴ cm²/(sec)(volt).

Convection or bulk flow. If the fluid as a whole is moving in the direction of diffusion with velocity V_z, component A is carried with it at a rate

$$N_{A_c} = V_z C_A \quad (5)$$

General equation for mass transfer flux. It is generally assumed that all mechanisms act simultaneously in an additive manner so that the equation for the net flux of a given component may be written as the sum of the fluxes due to each mechanism. For the flux of component A passing in the z direction through a unit cross section at any point within the fluid and at any particular instant of time

$$(N_A)_z = -D_A \frac{\partial C_A}{\partial z} - E \frac{\partial C_A}{\partial z} + D_{T\rho}' \frac{\partial \ln T}{\partial z} + C_A U_A \frac{\partial i}{\partial z} + V_z C_A \quad (6)$$

Similar equations can be written for the fluxes in other coordinate directions to give a set of general differential equations describing mass transfer in a fluid. These equations coupled with the equations for conservation of mass and energy offer the basis for general solution of mass transfer problems. Unfortunately, these equations seldom can be solved for situations of practical importance because knowledge of all coefficients, fluid velocities, and other variables must be known as a function of concentration and position.

In the absence of thermal diffusion and migration, Eq. (6) reduces to the diffusion-convection equation

$$(N_A)_z = -D_A \frac{\partial C_A}{\partial z} + V_z C_A \quad (7)$$

The diffusion-convection equations have been solved for a few situations such as the case of mass transfer from the wall of a tube to a fluid passing through it in viscous flow or for laminar flow over a flat plate. In these cases fluid velocities are known as a function of position in the fluid and the diffusion coefficient is constant so that the differential equations may be integrated.

Mass transfer coefficients. Mass transfer through a turbulent fluid to an interface formed by some other fluid with which it is immiscible or by a solid surface is commonly encountered in processes of industrial importance. For such situations it is necessary to resort to empirical equations to describe mass transfer rates. As an example, consider the dissolution of a slightly soluble solute such as benzoic acid from the wall of a pipe into water in turbulent flow. Solute concentrations in the fluid in the vicinity of the wall are illustrated qualitatively in Fig. 1. At the interface the water is saturated with solute at concentration C_i corresponding to the solubility of the solute. The concentration falls progressively from the wall until a value corresponding essentially to the value in the bulk liquid, C_0, is reached. Actually the concentration will ultimately reach some minimum value at the center of the pipe which will be very close to C_0 because of the flat concentration profile which exists in the turbulent core. The exact shape of the concentration profile will depend upon the magnitude of the molecular and eddy diffusion coefficients throughout the fluid. Very near the wall, where molecular diffusion is more nearly controlling, the profile will be steep and in the turbulent core, in which the eddy diffusivity is very large compared to the molecular value, the profile will be relatively flat. In the steady state the mass transfer flux at a given point is commonly expressed by a mass transfer coefficient, k_c, as follows:

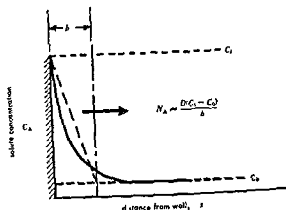


Fig. 1. Concentration profile for mass transfer from pipe wall into turbulent fluid.

$$N_A = k_c (C_i - C_0) \quad (8)$$

Similar concepts can be applied to both gases and liquids and for surfaces other than pipe walls. Various mass transfer coefficients may be defined corresponding to the units used for mass transfer rates, concentrations, and distance.

The most common concentration differences on which the various coefficients are based are partial pressure (for gases), molar concentration, and mole fraction. Thus for transfer in the gas phase of component A through a fluid to an interface

$$r_A = k_{gA} S(p_{A0} - p_{Ai}) = k_{YA} S(Y_{A0} - Y_{Ai}) = k_{CA} S(C_{A0} - C_{Ai}) \quad (9)$$

where r_A is the rate of transfer of component A in lb moles/hr, k_{gA} is the gas-phase mass transfer coefficient for component A in lb moles/(hr)(atm)(ft²), S is the interfacial area available for mass transfer in ft², p_{A0}, p_{Ai} are the partial pressure of

component A in the fluid bulk and interface respectively in atm k_{YA} is the gas-phase mass transfer coefficient mole fraction basis in lb moles/(hr)(ft²) Y_{A0} Y_{Ai} are the mole fractions of component A in the fluid bulk and interface respectively k_A is the gas phase mass transfer coefficient concentration basis in ft/hr and c_{A0} c_A are the concentration of A in the bulk and interface respectively

For liquid phase transfer similar equations may be written

$$r_A = k_{XA}S(X_{A0} - X_{Ai}) = k_{LA}S(C_{A0} - C_{Ai}) \quad (10)$$

where k_{XA} is the liquid phase mass transfer coefficient mole fraction basis lb mole/(hr)(ft²) X_{A0} X_{Ai} are the mole fraction of A in the fluid bulk and interface respectively and k_{LA} is the liquid phase mass transfer coefficient concentration basis in ft/hr Other symbols are the same as in the above equations for gases In many types of equipment fluid concentrations change over the mass transfer area and an appropriate average concentration difference between bulk fluid and interface must be employed

To illustrate the order of magnitude of mass transfer coefficients obtained under various conditions a range of typical values is listed in the table

Mass transfer coefficients in fluid systems

System (at 25°C 1 atm)	Gas phase water vapor to air stream k_Y lb moles/(hr)(ft²)	Liquid phase benzoic acid to water stream k_X lb moles/(hr)(ft²)
Pipe wall to fluid in 1 in i d tube at Reynolds number = 30 000	3 0	1 3
Pipe wall to fluid in 1 in i d tube at fluid velocity = 15 ft/sec	0 9?	4 2
Packed spheres to fluid ¼ in spheres at modified Reynolds number 1 000	6 2	3 5

Equivalent film thickness An alternate method of describing the rate of mass transfer is to define an equivalent film b of stagnant fluid which offers the same resistance to transfer (that is requires the same total drop in concentration between the interface and bulk liquid) as the actual fluid region which is not completely stagnant (Fig 1) Integration of Eq (1) for steady state diffusion gives the relation between the mass transfer flux and the film thickness for equivolume diffusion

$$N_A - \frac{D_A(C_i - C_0)}{b} \quad (11)$$

Correlation of mass transfer coefficients Theories for mass transfer in turbulent flow in pipes have yielded general correlations for the prediction of mass transfer coefficients *All of these depend* upon a knowledge of fluid velocity distribution and an assumption with respect to the

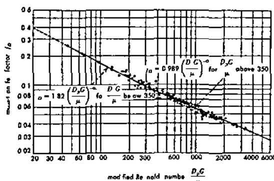


Fig 2 Mass transfer in the flow of fluids through granular beds (From C R Wilke and O A Hougen Estimation of liquid diffusion coefficients Chem Eng Progr 45 218–224 1949)

relationship between transfer of momentum and of mass Momentum transfer in turbulent flow can be described in terms of the interdiffusion of fluid eddies by means of an eddy diffusion coefficient for momentum analogous to the similar quantity for diffusion of mass One of the most successful of these semiempirical methods is that of C S Lin and associates in which the eddy diffusion coefficients for mass and momentum are assumed to be equal and to vary as the cube of the distance from the wall until the fully turbulent core is reached The resulting correlation has the following form

$$\frac{fU}{2k_c x_f} = 1 + \sqrt{\frac{f}{2}}\left[\frac{14\,5}{3}\left(\frac{\mu}{\rho D}\right)^{2/3} F\left(\frac{\mu}{\rho D}\right) + 5\ln\frac{1 + 5\,64(\mu/\rho D)}{6\,64[1 + 0\,41(\mu/\rho D)]} - 4\,77\right] \quad (12)$$

where

$$F\left(\frac{\mu}{\rho D}\right) = \frac{1}{2}\ln\frac{[1 + (5/14\,5)(\mu/\rho D)^{1/3}]^2}{1 - (5/14\,5)(\mu/\rho D)^{1/3} + (5/14\,5)(\mu/\rho D)^{2/3}} + \sqrt{3}\tan^{-1}\frac{(10/14\,5)(\mu/\rho D)^{1/3} - 1}{\sqrt{3}} + \frac{\pi\sqrt{3}}{6}$$

and x_f is a term for the effect of net fluid motion in the direction of mass transfer approximately unity for dilute solutions or for equivolume diffusion U is the average fluid velocity in the pipe f is the Fanning friction factor (a function of Reynolds number) μ is the fluid viscosity ρ is the fluid density and D is the diffusion coefficient of the component transferred

It is apparent that even for apparatus as simple as a pipe the relationship of variables is quite complex For turbulent flow in most systems the velocity distribution and other factors necessary for a similar analysis are unknown but it is generally assumed that the correlating variables for mass transfer appear as functions of the same dimensionless groups as for pipes A representative example is the correlation of the mass transfer factor j_d as a function of the modified Reynolds num

ber for transfer between solid particles in packed beds and fluids flowing through the beds as shown in Fig 2 The mass transfer factor and Reynolds number N_{Re} are defined as follows

$$j_d = \frac{k x_f}{U} \left(\frac{\mu}{\rho D}\right)^{2/3} \quad (13)$$

$$N_{Re} = \frac{D_p G}{\mu} \quad (14)$$

where U is the superficial fluid velocity through the bed $G = U\rho$ is the fluid mass velocity through the bed and D_p is the diameter of spherical particle having the same external surface area as the particles in the bed

Unsteady state diffusion and convection For unsteady state processes variation with time must be considered For diffusion and convection equations such as (7) may be written for each coordinate direction and combined with material balances over a fluid element during an interval of time $d\theta$ to give a partial differential equation which expresses the variation of concentration of a solute with time at a point within the fluid For an incompressible fluid with constant D_A

$$\frac{\partial C_A}{\partial \theta} = D_A \left[\frac{\partial^2 C_A}{\partial x^2} + \frac{\partial^2 C_A}{\partial y^2} + \frac{\partial^2 C_A}{\partial z^2}\right] - V_x \frac{\partial C_A}{\partial x} - V_y \frac{\partial C_A}{\partial y} - V_z \frac{\partial C_A}{\partial z} \quad (15)$$

where V_x, V_y and V_z are the components of fluid velocity in the coordinate directions x y and z respectively

Equation (15) may be combined with suitable relations describing the state of fluid motion to provide an analytical solution to mass transfer problems in a few cases involving fluids flowing in laminar motion If the fluid is at rest the last three terms on the right become zero and the problem becomes one of diffusion only which has been solved for many of the commonly encountered boundary conditions

Interdiffusion of two fluids As an example of unsteady state motion consider two pure fluids A and B initially separated by a diaphragm at the midpoint of a cylindrical container of total length L which are allowed to mix by diffusion For this case Eq (15) reduces to the following

$$\frac{\partial C_A}{\partial \theta} = D \frac{\partial^2 C_A}{\partial z^2} \quad (16)$$

This equation may be solved to give the mole fraction of fluid A in each half of the cylinder at time θ after removal of the diaphragm

$$Y_{A1} - Y_{A2} = \frac{8}{\pi^2} \left[\exp - \left(\frac{\pi}{L}\right)^2 D\theta + \frac{1}{9} \exp - 9\left(\frac{\pi}{L}\right)^2 D\theta + \frac{1}{25} \exp - 25\left(\frac{\pi}{L}\right)^2 D\theta + \cdots\right] \quad (17)$$

where Y [illegible] the cyl [illegible] mole fr [illegible]

[illegible] inally containing B and D is the diffusion coefficient for A in B This equation has been applied by A S Smith to the estimation of time required for mixing of gases in commercial cylinders of approximately 1 5 ft³ capacity, and 4 1 ft length at 25°C For example with butane air at 5 atm pressure $Y_{A1} - Y_{A2}$ is still 0 25 after 27 hours and [illegible]

exercised to assure complete mixing in the blending of gases Similar considerations apply for liquids in which diffusion times for equal degrees of mixing are much longer than for gases *See* MASS TRANSFER OPERATION [C R W]

Bibliography M Benedict and R L Pigford *Nuclear Chemical Engineering*, 1957, B W Gamson G Thodos and O A Hougen Heat, mass and momentum transfer in the flow of gases through granular solids *Trans Am Inst Chem Engrs*, 39 1–35 1943 C S Lin R W Moulton and G L [illegible]

[illegible] *Liquids* 1958 T K Sherwood and R L Pigford [illegible] *Ind* [illegible] *Mass* [illegible] A

Hougen Estimation of liquid diffusion coefficients *Chem Eng Progr*, 45 218–224 1949

Diffusion in solids

The term diffusion when applied to crystalline solids refers to the migration of atoms within a solid which results in actual mass transport through the crystal lattice associated with redistribution of the constituent atoms either unaccompanied by changes in chemical composition (self diffusion) or combined with such changes (chemical interdiffusion)

Diffusional motion is basic to many familiar reactions in solids In the photographic process the formation of the latent image and its development result from diffusion of silver atoms to initially irradiated points (*see* PHOTOGRAPHY) Electrical conduction in salts (ionic crystals) occurs by diffusional motion of ions under an electric field (*see* IONIC CRYSTALS) Metallurgy offers many examples of important diffusion limited reactions, such as welding sintering hardening recrystallization corrosion and creep

Diffusion mechanisms At temperatures well below the melting point diffusion occurs largely along grain boundaries and dislocations At higher temperatures where diffusion takes place through the bulk of the solid the exact mechanism is not known Four mechanisms which are illustrated in the figure have been considered (1) direct interchange—pairs or larger numbers of atoms interchange positions by rotation, (2) interstitial mechanism—motion of an atom directly between interstitial sites (interstitial) or by forcing a neighboring atom into an interstitial site (inter

stitialcy); (3) vacancy mechanism—jump of an atom into a vacant site; and (4) subboundary mechanism—atom motion along interconnecting low-angle boundaries.

In solid solutions in which one constituent is known to occupy interstitial lattice positions (for example, carbon in iron), diffusion apparently occurs by a simple interstitial mechanism. In substitutional systems best agreement between theory and experiment is generally found for the vacancy mechanism. In the interchange, interstitialcy, and vacancy mechanisms, motion of an impurity atom is correlated with motion of the surrounding atoms, whereas the interstitial and boundary models permit impurity motion in a rigid lattice.

Diffusion equations. Diffusional motion is described by the same equations that describe heat flow. See CONDUCTION (HEAT). The current of diffusing atoms per unit area, **J**, is given in terms of the composition, c (atoms/cm³), by Fick's first law,

$$\mathbf{J} = -D \operatorname{grad} c$$

where the diffusion coefficient, D, which has the dimensions of cm²/sec in cgs units, is proportional to the product of the frequency with which atoms jump between adjacent lattice sites and the square of the jump distance. The diffusion coefficient is a function of temperature, pressure, and chemical composition.

Because the total number of atoms remains constant during diffusion, the continuity condition requires that

$$\frac{\partial c}{\partial t} = -\operatorname{div} \mathbf{J}$$

Combining the equations, one obtains the diffusion equation, frequently designated as Fick's second law,

$$\frac{\partial c}{\partial t} = \operatorname{div} (D \operatorname{grad} c)$$

Solutions for constant D. For cases, as in self-diffusion, where the composition is stable and D is constant, the diffusion equation reduces to

$$\frac{\partial c}{\partial t} = D\nabla^2 c$$

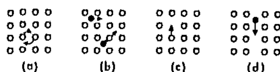
which can be integrated directly with the following error function solutions for typical cases:

Initial thin plane source $\quad c(x,t) = \dfrac{c(0,0)e^{-x^2/4Dt}}{(\pi Dt)^{1/2}}$

Initial point source $\quad c(r,t) = \dfrac{c(0,0)e^{-r^2/4Dt}}{8(\pi Dt)^{3/2}}$

Initial line source $\quad c(\rho,t) = \dfrac{c(0,0)e^{-\rho^2/4Dt}}{4Dt}$

In these equations x, r, and ρ are the appropriate coordinates in the directions of heat flow in the three cases. Thus x is the coordinate normal to the plane, r is the radial distance from the point source, and ρ that from the line source. The exponential dependence on distance is the same in all cases, and the mean square penetration depth is simply $4Dt$.



Four basic mechanisms for volume diffusion. (a) Direct interchange. (b) Interstitial (above) and interstitialcy (below). (c) Vacancy. (d) Subboundary.

Solutions for variable D. In general, because D is a function of the composition c and hence of distance and time, the diffusion equation is not separable and no analytic solution can be found. However, it is possible to determine a chemical interdiffusion coefficient, D_{chem}, as a function of composition from the experimentally measured variation of concentration with distance and time. For a one-dimensional case, the solution gives

$$D_{\text{chem}}(c_i) = \frac{1}{2t}\left(\frac{dx}{dc}\right)_{c_i} \int_{c_l}^{c_i} x\,dc$$

where the derivative and integral are evaluated numerically from the data for each composition c_i up to the limiting value c_l.

It can be shown that in a two-component system containing an atom fraction N_A of element A and N_B of element B, the chemical interdiffusion coefficient at the composition N_A is given in terms of the self-diffusion coefficients D_A and D_B of the two constituents measured in homogeneous specimens of composition N_A, by the relation

$$D_{\text{chem}} = (N_A D_B + N_B D_A)\left(1 + \frac{d \ln \gamma_A}{d \ln N_A}\right)$$

where γ_A is the thermodynamic activity of constituent A. In an ideal solution the derivative vanishes and the chemical interdiffusion coefficient is a weighted average of the self-diffusion coefficients of the two constituents.

Temperature dependence. The diffusion coefficient is found generally to vary exponentially with temperature as

$$D = D_0 e^{-Q/RT}$$

where the frequency factor D_0 and activation energy Q are independent of temperature, R is the gas constant, and T the absolute temperature. This relation has been justified theoretically by an approximate model which considers diffusion to be a simple isothermal rate process between atoms at equilibrium lattice positions and in so-called activated states midway between normal sites.

Under certain assumptions, D_0 is given approximately by the relation

$$D_0 = \alpha a^2 \nu e^{\Delta S/R}$$

in terms of a constant characteristic of the

symmetry α the jump distance a the Debye frequency ν and the change in entropy ΔS resulting from lattice strains during jump The activation energy is shown to be the change in enthalpy ΔH between equilibrium and activated states Calculated values for D_0 are in the range 0 01 10 cm² sec in excellent agreement with experiment

For cases in which the diffusion process involves a thermally generated imperfection of formation energy $\Delta E_f = \Delta H_f - T\Delta S_f$ where ΔS_f is the entropy of formation of the defect the measured Q is the sum of the enthalpies for motion ΔH_m and formation ΔH_f of the defect

Experimental methods Direct measurements of diffusion coefficients are best obtained where possible by use of radioactive tracers in extremely dilute solution consistent with the boundary conditions of the diffusion equation The value of D is determined directly from the tracer concentration by sectioning the sample after diffusion

Under some conditions it is more convenient to measure the diffusion coefficient indirectly In salts this can be done by measurements of electrical conductivity because the conductivity is directly proportional to D In other systems changes in nuclear magnetic resonance line widths and certain types of anelastic relaxation effects have been successfully correlated with diffusion These indirect methods are particularly applicable to studies at low temperatures and under nonequilibrium conditions

Experimental results Precise measurements of self-diffusion in various systems have shown a consistent correlation between measured activation energies for diffusion (Q) and melting temperatures the ratio of Q to melting temperature being roughly constant for many cases Impurities which lower the melting temperature generally diffuse faster than the matrix atoms and with lower activation energies On alloying with such impurities the rates of self diffusion of both solvent and solute atoms increase approximately exponentially with impurity concentration the effect being greater for the solvent Opposite results are found for impurities which raise the melting temperature Although these correlations are not well understood the effectiveness of an impurity in the diffusion process is apparently related to the effective chemical valence and ionic size of the impurity in solid solution

In chemical interdiffusion measurements in which inert markers are placed at the initial interface the markers are found to move with time in [illegible]

[illegible] terstitialcy mechanisms which permit changes in the total number of lattice sites in the diffusion zone In maintaining such a process in equilibrium dislocations and grain boundaries *apparently play* an important role as sources and sinks for the point imperfections Some imperfections also apparently precipitate about inclusions and produce marked porosity in the diffusion couple In the few cases for which data on both self diffusion and chemical interdiffusion are available, the results are in excellent agreement with the theoretical predictions See CRYSTAL DEFECTS [D L A]

Bibliography R M Barrer, *Diffusion in and Through Solids* 1951, *Defects in Crystalline Solids* Bristol Conf Rept, Physical Soc, 1955 W Shockley (ed) *Imperfections in Nearly Perfect Crystals* 1952

Digenea

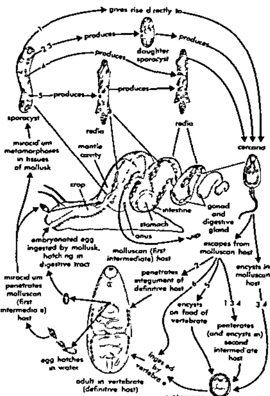
A group of parasitic flatworms or flukes constituting a subclass or order of the class Trematoda in the phylum Platyhelminthes The name Digenea refers to the two types of generations in the life cycle (1) the germinal sacs which parasitize the intermediate host (a mollusk or rarely an annelid) and reproduce asexually, and (2) the adult which is primarily endoparasitic in vertebrates and reproduces sexually The adult usually is hermaphroditic but many of the blood flukes and a few others are dioecious Vertebrates of all classes except the Cyclostomata serve as definitive hosts Those feeding on aquatic plants and animals harbor the greatest variety of digenetic trematodes but several species occur in strictly terrestrial hosts Effects on the vertebrate vary from no apparent harm to severe and [illegible]

Morphology The adult differs from other trematodes in several respects the most obvious being the absence of a posterior adhesive organ (opisthaptor) with sclerotized hooks plates or multiple suctorial structures Usually the mouth is anterior and opens into an oral sucker but in the gasterostomes it is ventral and without a sucker A second sucker the ventral one or acetabulum is at the posterior end of amphistomes on the ventral surface of distomes or absent in monostomes The entire ventral surface may be concave and serve for adhesion with or without the aid of a ventral sucker and glandular structures or a portion of that surface may form a complex adhesive organ as in the holostomes Adult digenetic trematodes vary from about 0 2 mm to well over a meter in length but usually are less than 1 in long They occur in any part of the vertebrate that provides egress for their eggs and also in the circulatory system from which eggs work through the tissues and escape in the feces or urine

Life histories of the digenetic trematodes These are complex and unlike those of other trematodes in which the egg gives rise to a single adult Instead the miracidium which develops from the egg enters the intermediate host and becomes a germinal sac the miracidial or mother sporocyst Its germinal cells give rise to a second or daughter sporocyst generation or to one of rediae which differ from sporocysts only in having a pharynx and saclike gut The redia which may or may not produce a second generation of rediae may be present or absent in the life history Germinal cells of the definitive sporocyst or redial generation develop

into cercariae (*see* CERCARIA) The miracidial sporocyst rarely produces cercariae directly and instances are known in which a redia or sporocyst may give rise to both its own kind and also to cercariae thus maintaining the infection indefinitely in the intermediate host after the production of cercariae begins The mode of reproduction in sporocysts and rediae has been disputed Most support is given to germinal lineage that is early segregation of germinal and somatic cells with polyembryony whereby cells of the germinal line multiply by mitosis often to form germinal masses and thus give rise to the next generation without maturation or other sexual phenomena The number of cercariae produced by the germinal sacs resulting from a single miracidium varies from a few dozen to thousands or even millions

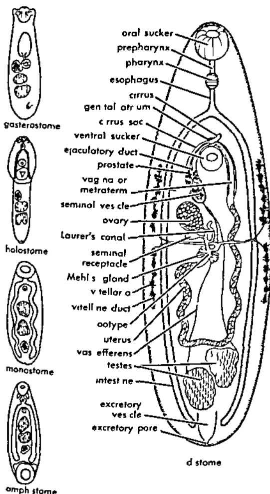
The cercaria usually leaves the intermediate host and then encysts either on vegetation or after penetrating a more or less specific second intermediate or vector host The encysted larva called the metacercaria is eaten by the vertebrate and becomes the adult fluke in that host Cercariae of blood



Some variations in life cycles of digenetic trematodes (*From R M Cable An Illustrated Laboratory Manual of Parasitology Burgess 1940*)

flukes enter the vertebrate directly by penetrating the skin Swimmer s itch in the United States results when such larvae of avian blood flukes penetrate the human skin and die because man is an unsuitable host Fish acquire certain trematodes by ingesting cercariae which are large and attract attention Cercariae rarely remain in the molluscan host with or without encysting and are eaten with the host by the vertebrate

Classification There is no general agreement concerning the classification of the Digenea When few life histories were known the group was divided into orders and families on the basis of such adult features as the number and position of suckers and the components and arrangement of the reproductive system Instances of divergent and convergent evolution with respect to those features have been revealed by life history studies which accordingly have prompted drastic revision of taxonomy of the Digenea especially at the superfamily and order levels The scheme of G La Rue proposes two superorders the Anepitheliocystida and Epitheliocystida based on the presence or absence of an epithelial lining of the excretory bladder as revealed by studies on cercariae Each superorder is divided into orders on the basis of the cercarial tail the relationship of the excretory system to it and the structure of the miracidium The cercarial type is the basis of the superfamily with that category being divided into families largely according to adult structure In that respect, trematodes of the



Digenetic trematodes Diagrams of various adult types (*From R M Cable An Illustrated Laboratory Manual of Parasitology Burgess 1940*)

subclass or order Aspidobothria (Aspidogastrea) resemble the Digenea more than the Monogenea but so far as is known do not have germinal sacs and larval reproduction in an intermediate host *See* ASPIDOGASTREA PLATYHELMINTHES [R M C]

Digestive gland

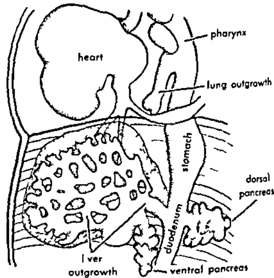
One of many structures found associated with the digestive tract many of which are both exocrine and endocrine They function primarily in digestion Like skin glands the glands of the digestive tract fall into two main groups those which arise and continue to reside within the epithelial layer and those which evaginate during development into subepithelial areas Although derived from oral ectoderm salivary or oral glands function during digestion (*see* ORAL GLAND) Glands which arise in situ within the epithelium are mucus secreting and are either single unicellular glands such as the goblet cells in the intestines or multicellular epithelial sheets present as surface cells of the gastric mucosa Among these glands are those of the pharynx esophagus stomach intestine and such structures as the liver and pancreas whose exocrine secretions empty into the small intestine

Pharynx and esophagus Glands of the pharynx are branched tubular mucus secreting glands which in development have pushed deeply into the subepithelial areas Glands of the esophagus are predominantly compound tubuloacinous structures secreting mucus Compound tubular cardiac glands similar to those in the esophageal end of the stomach also are present in the end of the esophagus near the stomach These cardiac glands may be found scattered in other areas of the esophagus In some mammals such as the cat and rat glands are absent in the esophagus

Stomach The most important digestive glands of the stomach are simple branched tubular structures They develop as a slender evagination of the lining epithelium into the submucosal areas They are in the form of mucus secreting sheets in the ducts of the glands and upon the surface epithelium of the stomach These glands secrete gastric juice composed of mucus digestive ferments and hydrochloric acid Cardiac glands near the esophagus are compound tubular structures capable of secreting digestive enzymes

Intestine The intestine contains the glands of Brunner or duodenal glands crypts of Lieberkuhn also known as the intestinal glands and the associated organs the liver and pancreas The crypts of Lieberkuhn are numerous simple tubular glands which arise as evaginations into mucosa of the small intestine The glands of Brunner in the duodenum are simple branched tubular glands They evaginate from the bottom of the crypts of Lieberkuhn and push into the submucosa *See* DIGESTIVE SYSTEM

The liver is the largest gland of the body It is a compound tubular gland having both exocrine and humoral secreting cells The liver arises as a ventral evagination from the entodermal epithelium of the duodenal area of the gut immediately posterior to the stomach The original evagination branches and rebranches to form a complex system of tubules which push into and involve the substance of the ventral mesentery in lower vertebrates or the mesenchyme of the septum transversum in higher vertebrates Much of the liver substance is derived from blood vessels and tissue which comes to surround the outgrowing ducts and tubules *See* LIVER



Early development of the digestive glands showing liver and pancreas

The pancreas is a compound acinous gland with exocrine and endocrine functions (see illustration) In mammals it arises by two distinct outgrowths or evaginations of the entodermal epithelium associated with liver outgrowth One outgrowth is ventral the other dorsal and they later fuse into one elongated mass The main pancreatic duct is formed from the fusion of part of the duct of the dorsal pancreatic outgrowth with part of the duct from the ventral pancreatic outgrowth The accessory pancreatic duct is derived from the remainder of the duct from the dorsal pancreatic outgrowth The dorsal and ventral pancreatic outgrowths at first form complex tubules Later distal areas of these tubules develop acinous outgrowths Endocrine cells or pancreatic islets are represented by large numbers of isolated masses the islands of Langerhans which probably develop from the distal ends of the tubules and hence are not modified acini *See* PANCREAS [O F N]

Digestive system

That system of structures in which food substances are digested that is complex food materials are subjected to the action of digestive enzymes throughout the system in the presence of water and broken down into simpler chemical compounds Absorption into the cells and tissues is then possible The digestive system of invertebrates varies among the various phyla Many protozoans have no permanent structures such animals as sponges and coelenterates morphologically have an incom

plete digestive system whereas most animals have a complete dige tive system in that there is a mouth and anal opening Among the vertebrate classes the structures of the digestive system are essentially similar although they may be highly modified ac cording to the food habits of the animal (*see* FEED ING MECHANISMS AND DIGESTION) This article covers the embryology anatomy and histology and physiology of the digestive system

EMBRYOLOGY

Development of the digestive tract The ferti lized eggs of vertebrates begin their development with the process of cleavage by which the whole egg (cyclostomes amphibians mammals) or a small plate of living tissue resting on a large yolk (fishes reptiles birds) is converted into a mass of small cells (*see* EMBRYOLOGY) These cells are soon resolved into three tissue sheets the germ layers The innermost of these layers the endoderm is the source material of the essential tissue of the digestive tract and most of its associated structures The initial disposition of the endoderm varies from class to class In the amphibians the endoderm constitutes a tube or trough as soon as it takes up its proper position within the embryonic body In the large yolked types the endoderm begins as a flat sheet which is converted into a tube by folding under at the tip and along the sides of the future embryonic body In mammals although the egg is virtually without yolk the endoderm follows the pattern laid down in the large yolked ancestors At an early stage in all embryos the endoderm forms an intimate union with a sheet of mesoderm which is the middle germ layer This mesoderm later differentiates into the musculature and sup porting and binding tissue of the digestive organs

When the endoderm tube is definitively formed it consists of three major regions the foregut or pharynx midgut and hindgut The foregut at its anterior end touches on the ectoderm which makes up the outer surface of the body and the hindgut makes a similar contact When these points of con tact break through the gut has established its communications with the external environment In embryos which have a large quantity of yolk in a sac partially separated from the developing body the midgut is open on its lower (ventral) surface into the yolk the same arrangement occurs in mammalian embryos despite the absence of actual yolk The connection between gut and yolk sac be comes progressively smaller but is not entirely eliminated until the fetus is born or hatched There is however no actual passage of food material directly from the yolk sac into the developing digestive tract In reptiles and birds the yolk is digested by enzymes produced by the yolk sac it self the small molecules thus produced being car ried into the embryonic body by way of the circula tory system

At the place where the foregut touches the ecto derm there is formed on the body surface a de pression known as the stomodeum This small cavity is the forerunner of the mouth which is therefore lined not with endoderm which lines the rest of the digestive tract but with the ectoderm that else where gives rise to the epidermis This fact is of importance in evolution because skin has the capacity to form bony scales and teeth have ap parently evolved from modified scales that ap peared in the mouths of ancient fishes The salivary glands are also formed from mouth ectoderm and arise by outpocketing into the mesodermal tissues of the jaws and head At the end of the hindgut there is a similar ectoderm lined depression the proctodeum which becomes incorporated into the anus *See* ORAL GLAND TOOTH

Foregut Early in development the foregut or pharynx is the largest and most complex part of the digestive tube The embryonic pharynx is a wide space its lateral walls being pulled out at five or more intervals in extensions that touch the

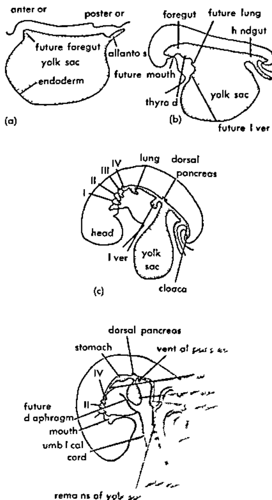


Fig 1 Early stages ... of human embryo (a ... (thy) resp ratory ... (c) About 25 days ... pouches (d) About ... parat ve Embry... McGraw H ll 19...

outer ectoderm The presence of these extensions the visceral pouches breaks up the thick lateral walls of the body into a series of mesodermal masses called visceral arches In a fish embryo the pharynx persists the visceral pouches give way to the gill slits and the visceral arch masses become elaborated into gills In terrestrial forms the predominance of the embryonic pharynx is transient but it produces a number of important derivatives (1) The middle ear cavities are formed as upward extensions from the first pair of pouches the persistent connections between these extensions and the back of the mouth become the Eustachian tubes (2) The two pairs of parathyroid glands arise as thickenings on the surfaces of the third and fourth pouches (3) The thymus gland begins as ventral extensions from the third pouches the two extensions later fuse to form a single gland (4) The thyroid develops as an outpocketing from the floor of the pharynx When the growth rate of the pharynx slackens the glandular derivatives split away from it and are carried by movements of the growing body to their definitive locations the thyroid and parathyroids in the neck and the thymus in the chest This shifting process is underway at the beginning of the third month in the human embryo The pharynx itself is finally reduced to a small area where the digestive and respiratory tracts cross Just posterior to the pharynx the endodermal tube gives rise to another downward *diverticulum* from which the trachea and lungs are formed *See* PHARYNX RESPIRATORY SYSTEM

Early stages in the de [illegible]
tract of [illegible]
of the y [illegible]
begins to produce a foregut Two days later infolding of both the foregut and hindgut is well advanced The rudiments of the *thyroid respiratory* tract and liver begin to appear in the anterior area [illegible]
closed [illegible]
openin [illegible]
days the major subdivisions of the digestive tract can be identified in the embryo which is about 5 mm long

Digestive tube proper For a short time the digestive tract posterior to the pharynx remains a simple tube Very soon however (the beginning of the second month in the human *embryo*) some differentiation becomes apparent as the region of the future stomach widens and the future intestine grows longer in mammals it sends a long loop into the base of the umbilical *cord* At its terminal end the tube widens into a chamber called the cloaca which also receives the mesonephric ducts from the embryonic kidneys *The cloaca* makes contact with the body surface at a point where the anal opening is later formed *See* URINARY SYSTEM

The widening of the stomach proceeds rapidly along the upper (dorsal) *surface of the organ* As the expansion goes on the organ rotates through 90° around its long axis and at the same time pulls away from its original position coming to lie across the long axis of the body rather than parallel with it (human 7 weeks) Normally the incurrent (cardiac) end of the stomach is on the left the excurrent (pyloric) end on the right This process is quite uniform among vertebrates, except for those that have a greatly elongated body form such as snakes and for birds, in which the stomach region becomes subdivided into two parts one specializing in the production of digestive enzymes (proventriculus) and the other in mechanical grinding (gizzard) As the stomach assumes its definitive form the intestinal loop undergoes torsion to form a rapidly growing coil that at first lies within the umbilical cord but later (human, 10 weeks) is withdrawn with the result that the small intestine appears in its normal position The large intestine grows more slowly and does not attain its characteristic form (ascending transverse and descending portions) until the latter half of gestation In mammals except the very primitive monotremes the cloaca is subdivided at an early stage into digestive and urogenital portions, the digestive part constitutes the rectum which opens independently to the body surface at the anus In all other vertebrates the cloaca persists as a common chamber entered by the large intestine, the kidney ducts and the genital ducts

The mesodermal sheet that associates itself with the endoderm at the onset of body formation contributes importantly to the walls of the digestive tract By 6 weeks in the human embryo this sheet proliferates a thick layer of loose cells that later differentiate into the spongy mucosa and the two muscular layers of the digestive organs The original sheet remains on the outer surface and becomes a layer of tough connective tissue (serosa) which covers the organs and binds nerves and blood vessels onto their surfaces Because the inner layers grow faster than the outer, the mucosa and its covering endoderm are thrown into folds which are shallow in the stomach but deeper in parts of the intestine Growth is so rapid that the lumen of much of the digestive tract is for a time occluded by the accumulation of large numbers of endodermal cells, subsequently the cavity is reestablished In the intestine, the folded mucosal tissue becomes covered by enormous numbers of microscopic processes called villi, these begin to appear in the human embryo at 8 weeks The villi tremendously increase the absorptive surface of the intestine The original endodermal sheet in the stomach and intestine differentiates into a single cell layer of tall narrow cells that cover the mucosal surface These cells develop the biochemical equipment that enables them to produce enzymes and accessory substances such as hydrochloric acid and mucus In the intestinal region they also acquire the ability to absorb digested food molecules and pass them into blood vessels in the intestinal wall In the esophagus which is the passageway of the digestive tract through the chest region the endoderm comes to consist of many layers of flattened cells

Liver and pancreas. The liver appears early (human 3½ weeks) as an outpocketing of the floor of the endodermal tube at the posterior end of the future stomach The pancreas develops in a similar manner a few days later The point of origin of the liver bud persists as the common bile duct connecting the liver to the small intestine In its early development the bud subdivides into five parts each served by a hepatic duct that leads into the common bile duct Another diverticulum from the liver bud gives rise to the gallbladder and its cystic duct

The early subdivisions of the liver bud rapidly send out cords of cells the liver cords which form a spongework that surrounds and invades two large veins leading into the heart immediately in front of the liver Thus the liver directly acquires its rich bed of fine blood spaces (sinusoids) Growth is so intense that by 7 weeks in the human the liver is larger than the heart although the heart is functioning before the liver bud appears The cords of cells come to enclose fine tubules called bile capillaries These tubules collect the bile which is manufactured by the liver cells and convey it to the hepatic ducts When digestion is not proceeding the bile is stored in the gallbladder a thin walled sac that becomes embedded in the liver *See* CARDIOVASCULAR SYSTEM

The liver is peculiar because it receives blood not only from the arterial system but also from the venous system by way of the hepatic portal vein This portal system is formed because the veins that are invaded by the rapidly expanding liver mass are connected early in development with a tributary from the subintestinal area In time this tributary becomes joined by a system of veins draining the stomach intestines pancreas and spleen Thus the venous blood from the viscera is carried to the liver where it passes through the bed of sinusoids before being collected into a vein again to be carried to the heart This arrangement permits the liver to remove sugar and other substances from the blood when digestion is proceeding actively *See* GALLBLADDER, LIVER

The pancreas first appears in mammals as two independent outpocketings one from the floor of the intestine just posterior to the point of origin of the liver diverticulum the other from the roof of the intestine, in other vertebrates there may be three outpocketings The two diverticula of mammals develop separately for some time but later come together to form a single organ in which two distinct lobes may be recognized When fusion occurs the original connections into the small intestine are both retained but only one develops functionally the other dwindles to a vestige In the human fetus the ventral connection becomes the actual pancreatic duct the other constitutes the usually nonfunctional accessory duct of Santorini In other mammals the situation may be reversed

Both pancreatic buds develop by repeated bifurcations so that the central duct acquires a large number of ever finer branches that termin microscopic sacs called acini These acini composed of a single layer of large cells produce the digestive enzymes that make up the pancreatic juice This juice is collected by the duct system and poured into the small intestine by way of the pancreatic duct

In addition to forming the acini and ducts the dorsal lobe of the developing pancreas also buds off separated cells that organize themselves into the islets of Langerhans These islets which comprise a large part of the fully developed pancreas constitute the endocrine portion of this organ The islet cells acquire the ability to synthesize insulin and glucagon which they release directly into the blood vessels

In contrast to the stomach and intestines in which only the single cell lining layer is of endodermal origin the liver and pancreas are largely endodermal although recent studies indicate that a variable part of the proper tissue of the liver may be derived from loose mesoderm Both organs are bound on their outer surfaces by tough coats of serosa *See* ENDOCRINE GLAND PANCREAS

Suspension of the digestive organs As the splanchnopleure forms itself into a tube the covering mesoderm is continuous with the mesodermal sheet that lines the surface of the abdominal cavity On the lower side this connection quickly disappears except in the liver region On the upper side the connection persists generally as the dorsal mesentery an extensive sheet of thin membranous tissue that attaches the digestive organs to the mass of musculature surrounding the backbone Thus the stomach and intestines are suspended from above but are free to move on the ventral side In the human being apparently as an adaptation to erect posture the freedom of the intestinal coils to move about is considerably reduced by fusions to the

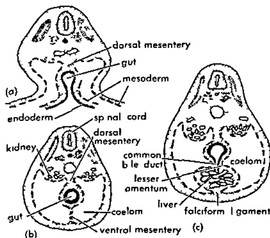


Fig 2 Development of principal mesenteries of the gut in higher vertebrates (a) Closure of the gut on its lower side by approximation of folds of splanchnopleure (b) Gut and body wall closed by sealing together of ventral folds (c) Persistence of ventral mes

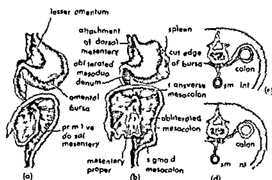


Fig 3 Rotation of the gut and secondary fusions of mesenteries of duodenum and colon (a) At 3 months *before fusions begin (b) Later stage fused surfaces indicated by cross hatching (c d) Method of mesenteric obliteration (transverse section) (From L B Arey Developmental Anatomy 6th ed Saunders 1954)*

muscular mass of the back The duodenum becomes fused at about 11 weeks and later the ascending and descending limbs of the colon are fixed in place Only the coils of the small intestine retain the mesentery and continue to be free to move within the restricted coelomic space In mammals (Fig 3) the mesentery of the stomach (greater omentum) expands greatly during and after rotation of the stomach so that it constitutes a large double apron of thin membranous tissue lying over the intestines it serves among other functions as a fat depository The esophagus is from the first embedded in the dorsal musculature in the chest region and consequently never develops a supporting mesentery Where the liver arises near the gastric intestinal junction, the ventral mesodermal connections persist and become elaborated into the lesser omentum which attaches the liver to the stomach and duodenum above and the falciform ligament which attaches the liver to the body wall below and to the diaphragm or transverse septum in front

Development of function The capacity for function in the digestive tract is largely developed before birth The liver which has been most fully investigated from the point of view of functional differentiation deposits glycogen at 7 days in the chick embryo and at a comparably early time in mammalian embryos At about the same time the organ begins to manufacture and secrete bile and in mammals it is active in embryonic and fetal stages in the production of blood cells a function later assumed by the bone marrow The pancreas begins to elaborate its proteolytic secretion near the end of fetal life, but appears to produce its endocrine products somewhat earlier Swallowing is possible relatively early in gestation (human 4-5 months) and propulsive movements of the stomach and intestines occur at the same time Thus a part of the amniotic fluid in which the embryo is bathed may be in continuous passage through the intestinal tract although the extent to which these events occur under normal conditions is not known X ray studies of passage of radiopaque material through the gut of exteriorized fetuses show that the peristaltic contractions become more vigorous as the fetus grows older Dyes injected into the amniotic fluid can be absorbed through the intestinal wall but there is no evidence that normal digestion and absorption go on even though the gut accumulates a considerable quantity of desquamated lining cells and other substances In bird embryos there is evidence that some of the albumen is swallowed about midway through the period of incubation This accumulated material is the source of the meconium that is usually voided shortly after birth or hatching According to recent studies, the intestinal epithelium of the chick embryo acquires many of the specific structural and chemical features characteristic of the functional state only shortly before hatching, in the mouse these changes occur partly before birth, but partly also at the time of weaning The nursling mouse when first able to ingest the adult diet is not able to digest or absorb it Hence it seems unlikely that normal digestive and absorptive capacity exists during fetal stages

Causal relations. The development of organs from layers of undifferentiated tissue poses an important question concerning the origin of the organs whether the organ forming property is inherent in the undifferentiated tissue or whether it arises through interactions in the course of development (see EMBRYOLOGY, EXPERIMENTAL) If the splanchnopleure (that is, endoderm and its associated mesoderm) of a very early chick embryo is cut into pieces and transplanted to foreign sites, histologically identifiable liver, intestine and also thyroid appear, stomach and esophagus appear too, although they are less clearly recognizable The distribution of these experimentally realized potencies in the undifferentiated splanchnopleure corresponds roughly with the areas that would normally give rise to the same organs Similar results have been obtained in amphibians When liver tissue appears in grafts, however, heart tissue is usually found too and there is reason to believe that contact with the heart forming area is essential to liver development This relationship seems to be involved in the causation of situs inversus because this side for side transposition of stomach liver and intestines can be caused experimentally by injuries inflicted on the mesoderm particularly on the left side where the stronger heart forming potency resides The endoderm itself exerts inductive influences on tissues that are associated with it In particular, the normal formation of the mouth and the perforation of the anus depend on the establishment of contact between endoderm and the overlying ectoderm Gills fail to develop in amphibian embryos if the gill buds are deprived of their component of foregut endoderm If the entire endoderm is extirpated from early amphibian embryos, the larva that develops shows defects that cannot be ascribed solely to lack of the mechanical

support that the yolk laden endoderm provides Two pairs of forelimbs frequently develop the heart fails to appear, and the head may show serious abnormalities such as the development of a single median eye *See* EMBRYONIC INDUCTION

Anomalies Defects in the digestive tract of the newborn animal are not uncommon, because failure of any one of the complex series of movements, interactions, and other differentiative events involved in the complete development of the tract will lead to manifest abnormality Narrowing (stenosis) or complete absence (atresia) of the lumen of the esophagus, stomach, or intestines apparently results from persistence of an occluded condition that is a transient stage in gut development The entire digestive tract may be transposed from left to right, or only the intestines may be reversed in position, the latter condition probably reflects a disturbance during withdrawal of the intestine from the umbilical cord A more serious condition is umbilical hernia, in which the withdrawal process fails and some or all of the loops of the small intestine protrude outside the abdominal wall Occasionally a connection (Meckel's diverticulum) persists between the lower part of the small intestine and the umbilicus marking the original attachment of the yolk sac to the midgut Failure of the endoderm to break through to the body surface results in an imperforate anus, which may be accompanied by atresia of the rectum Sometimes the rectum fails to split away from the urogenital chamber in an embryonic mammal so that a reptilian type of cloaca persists In viable fetuses the liver is not subject to abnormalities affecting its function, but the number of lobes may be greater or smaller than normal, and the gallbladder may fail to appear altogether Accessory pancreases are a common result of the appearance of supernumerary primordia, but occasionally only one primordium develops [F M]

ANATOMY AND HISTOLOGY

The digestive tract proper (or gut) consists of an elongated muscular tube lined with a mucous membrane and extending from the mouth to the anus The muscular wall serves to propel and mix its contents while the epithelium on its inner surface acts as a protective barrier as well as a source of digestive juices and in the absorption of the products of digestion Because the lumen of the gut is in communication with the outside of the body, its contents cannot be considered truly to have entered the body until absorption has occurred

Four successive segments of the gut are distinguished on the basis of their structural and functional specializations (Fig 4) the esophagus for rapid conveyance of food from the mouth to the stomach, the stomach where the food is held for a time and digestion begins, the small intestine where digestion and absorption are virtually completed, and the large intestine into which undigested and waste materials are passed for elimination by way of the rectum and anus

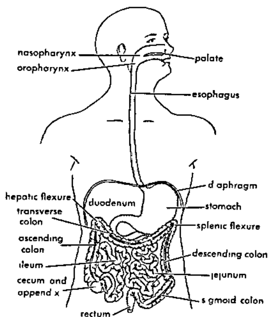


Fig 4 Digestive tract in man (*From L L Langley, E Cheraskin, and R Sleeper, Dynamic Anatomy and Physiology McGraw Hill 1958*)

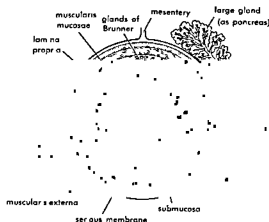


Fig 5 Diagrammatic cross section of the intestinal tract In the upper half of the drawing the mucous membrane is provided with glands and villi in the lower half it contains only glands (*From A A Maximow and W Bloom, A Textbook of Histology, 7th ed, Saunders, 1957*)

General histology. Although marked variations occur in its different segments and in various vertebrate groups the wall of the digestive tube can in general be described in four successive layers (Fig 5) Innermost is the mucous membrane (mucosa) which consists of an epithelial lining supported by a layer of delicate interlacing connective tissue fibers, the lamina propria The outer limit of the mucosa is generally marked by a thin layer of smooth muscle fibers, the muscularis mucosae The submucosa, lying beneath the mucosa, is composed of loosely interwoven somewhat

coarser connective tissue fibers and serves as attachment for the mucosa while at the same time allowing considerable freedom of motion The muscularis externa consists of smooth muscle with fibers generally arranged as a distinct inner circular and an outer longitudinal layer That part of the gut which lies free within the abdominal cavity is covered externally with a serosa which consists of a thin layer of connective tissue with peritoneal epithelium giving it a smooth outer surface The serosa is continuous with the mesentery attaching the intestine to the abdominal wall and thence with the peritoneal lining of the abdominal cavity

Blood vessels and nerves reach the gut wall by way of the mesentery Branching vessels penetrate into the lamina propria where they form a rich capillary network beneath the epithelium

Evaginations from the epithelium form glandular structures with varying degrees of complexity The simplest of these are confined to the lamina propria others penetrate into the submucosa The pancreas and liver remain connected with the intestine only through elongated ducts Networks of lymph vessels in the lamina propria and submucosa play an important role in the transport of absorbed food substances

Nerves of the autonomic (involuntary) nervous system enter the gut wall and form a netlike plexus of fibers between the layers of the muscularis externa (the myenteric plexus of Auerbach) A second plexus (Meissner s) is formed within the submucosa Nerve cells and fibers associated with these plexuses regulate blood flow glandular secretion and gastrointestinal movement Sensory nerve fibers within the gut wall also play a role in reflex regulation of the activities of the digestive tract

Esophagus The esophagus is a strong muscular tube for the rapid peristaltic transport of food from pharynx to stomach In fishes and amphibians (Fig 7*a*) it is generally short and ill defined (Fig 6*b e f*) although in a few primitive vertebrates which lack a stomach (cyclostomes chimaeras) the entire segment between pharynx and small intestine may be considered esophagus (Fig 6*a c d*) The swim bladder of fishes develops as an outgrowth from the esophagus with which it may remain connected through a pneumatic duct In terrestrial vertebrates with reduction in size of the pharyngeal region and development of a neck and thorax the esophagus becomes a distinctive structure In man it is about 10 in long lies close to the vertebrae as it traverses the thorax and pierces the diaphragm just before joining the stomach *See* ESOPHAGUS SWIMBLADDER

The crop (ingluvies) occurs in birds especially birds of prey and seed eaters as a distensible saccular diverticulum near the lower end of the esophagus (Figs 7*c* 8) This serves as a temporary holding station and permits the rapid ingestion of large amounts of food Although digestion apparently does not occur in the crop food there becomes moistened and softened by salivary and esophageal secretions and is thus rendered more

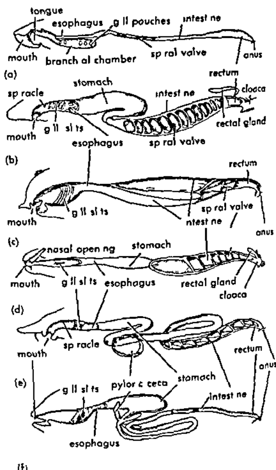


Fig 6 Diagrams of digestive tract in (*a*) lamprey (*b*) shark (*c*) chimaera (*d*) lungfish (*e*) sturgeon (*f*) teleost (perch) The stomach of the lungfish is nonglandular and is simply an enlarged segment of the esophagus (*From A S Romer The Vertebrate Body Saunders 1949*)

easily digestible when passed on into the stomach

In pigeons the epithelial lining of the crop is specialized to form the so-called crop glands In both sexes during the breeding season epithelial cells in these areas increase in number become laden with fatty material and are shed into the lumen of the crop This curdlike substance known as pigeon s milk is regurgitated as food for the young However it is quite different from mammalian milk because it contains neither casein nor lactose Production of this material is controlled hormonally and can be induced experimentally by injection of the pituitary hormone prolactin

Histology The epithelium lining the esophagus in fishes generally consists of several layers of flattened cells often interspersed with mucus secreting cells In amphibians and reptiles the surface epithelium is usually ciliated with the ciliary current directed toward the stomach Some amphibians have a highly vascularized esophageal epithelium the surface here (and in the mouth) plays a role in ex In birds and mamm

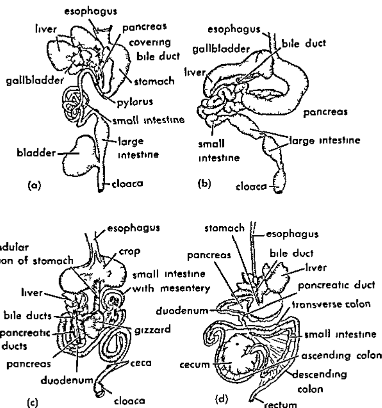


Fig 7 Diagrams of digestive tract and appendages (ventral view) in (a) frog (b) reptile (horned toad) (c) bird (pigeon) (d) mammal (guinea pig) (*After Gaupp (a) Potter (b) and Schimkewitsch (c) in A S Romer The Vertebrate Body Saunders 1949*)

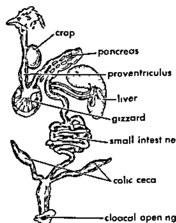
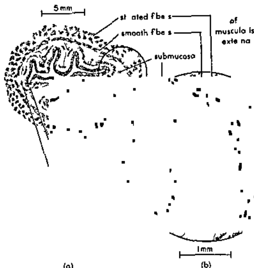


Fig 8 Digestive tract of a hen (*From H E Walter and L P Sayles Biology of the Vertebrates 3d ed Macm llan 1949*)

stratified squamous epithelium which provides a tough and often cornified surface layer well suited to withstand the abrasive action of rough food in transit (Fig 9) Numerous horny papillae di rected toward the stomach are present in marine turtles and some birds but in mammals the surface is smooth and slippery In the quiescent esophagus the entire mucous membrane including the muscularis mucosae is thrown passively into numerous longitudinal folds which practically obliterate the lumen These folds are smoothed out when the lumen is distended during the passage of a bolus of food *See* EPITHELIUM

The muscularis externa is composed of an inner circular and outer longitudinal layer of smooth (involuntary) muscle Between these is located the myenteric plexus (Auerbach's) of the autonomic nervous system In most mammals including man the muscle in the upper quarter of the esophagus is of the striated type and under voluntary control at least so far as participating in the act of swallow ing is concerned Below this level striated muscle gradually gives way to smooth muscle In the

Glands of the esophagus are of two types the esophageal glands proper lying within the sub mucosa and the smaller cardiac glands confined to the lamina propria Both develop as epithelial invaginations from the lining of the esophagus and discharge their secretion (chiefly mucus) through ducts opening into the lumen of the esophagus These secretions, together with the saliva serve to lubricate the esophagus There is great variation the number and distribution of such glands



F g 9 Cross sect on of the esophagus (a) Low power wall (b) H gh power (From A W Ham H stology 3d ed L pp ncott 1957)

birds the glands are confined to the lamina propria and often are located almost entirely within the epithelium In the dog esophageal glands are very numerous and even extend for a short distance into the wall of the stomach *See* DIGESTIVE GLAND

The connective tissue immediately surrounding the esophagus as it traverses the neck and thorax is known as the adventitia

Blood supply The blood supply of the esophagus in mammals is derived from numerous small arterial branches along its course which include the inferior thyroid and bronchial arteries in the neck intercostals in the thorax and the diaphragmatic and left gastric arteries in the abdomen Veins draining the esophagus establish at its lower end anastomotic connections with gastric veins draining into the hepatic portal system These connections may become an important channel for the return of venous blood from the gastrointestinal tract in pathological conditions which obstruct the portal vein

Nerve supply The striated muscle of the upper esophagus (entire length in ruminants) is innervated by voluntary motor fibers of the vagus The smooth muscle of the esophagus is innervated by parasympathetic fibers of the vagus These nerves run parallel to the esophagus and send branches which penetrate to join the myenteric and submucous plexuses

Stomach The stomach is a pouchlike dilatation of the gut immediately beyond the esophagus in which food is held for a time and subjected to vigorous digestive action

Fishes The stomach of fishes is generally simple and saccular in contour Some are more or less straight others are curved or J shaped (Fig 6) Although most stomachs have a digestive function there are a few (lungfishes) in which the mucous membrane produces no digestive juices Among both fishes and amphibians patches of the surface epithelium may be ciliated Entrance into the stomach is controlled by a cardiac sphincter which

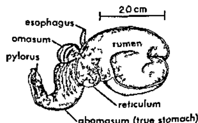
d e entrance of water

variations in stomach configuration other than those required to fit it within the available space In crocodiles the anterior part of the stomach is enlarged and very muscular and has a structure resembling the gizzard of birds

Birds The stomach of birds is divided into two chambers the relatively thin walled glandular proventriculus and the very muscular gizzard (Figs 7c 8) Within the proventriculus the food is mixed with mucus and digestive juices produced by the glandular mucous membrane The softened and partially digested mass is then passed on into the gizzard where it is vigorously ground and comminuted as digestion continues Grinding action of the gizzard highly developed in seed eating birds is augmented by the presence of gizzard stones which these birds ingest Digestive juices are not produced within the gizzard but its richly glandular mucosa produces a secretion which hardens into a tough horny surface cuticle well adapted to withstand the abrasive and digestive action The muscular wall of the gizzard is formed by a great enlargement and specialized arrangement of an inner oblique and middle circular layer of smooth muscle The outer longitudinal layer is much reduced or absent On one or both sides of the gizzard the wall generally has a dense fibrous mass from which the muscle bundles radiate Immediately beyond the gizzard there is generally a short prepyloric segment

Ruminants The most elaborate mammalian stomachs are found among the cud chewing artiodactyl (even toed) ungulates such as the cow sheep deer goat giraffe and camel Stomachs of the ruminants are not only complex but large, that of an ox has a capacity of up to 60 gallons Typically the stomach is divided into four chambers (Fig 10) rumen (paunch) reticulum (honeycomb) omasum (psalterium or manyplies) and abomasum The first three are generally considered to be expansions of

mals to ingest large amounts of food in a rela



F g 10 Stomach of a sheep opened to show the four chambers characteristic of h gher ruminants (After Pernkopf in A S Romer The Vertebrate Body Saunders 1949)

tively short period of time and then rework and digest it

The rumen, the largest of the chambers (about 80% of the total volume) serves as a receiving space for ingested herbage Within the rumen food mixed with saliva is churned about in a rotary motion and undergoes some fermentation and bacterial digestion It then passes gradually through a rather large orifice into the reticulum Here the pulpy mass is formed into cuds which can be regurgitated voluntarily by action of the striated esophageal muscle which extends onto the stomach After thorough chewing and mixing with saliva the now finely divided solids and juices are again swallowed this time following a groove which bypasses the rumen and reticulum and leads into the omasum Here the contents are mixed to a more or less homogeneous state and passed on into the abomasum Within this final chamber mucus and digestive juices are added from the glandular mucosa and true gastric digestion occurs The contents of the abomasum are gradually passed through the pylorus into the duodenum

A stratified squamous epithelium without glands lines the first three chambers in the cow for example (Fig 12) The inner surface of the rumen is studded with many small horny papillae In the reticulum the mucous membrane is thrown into numerous coarse and fine ridges which outline variously shaped areas or "cells in an arrangement which has been called the honeycomb The mucosa of the omasum (psalterium) is elevated into many thin crescent shaped folds which suggested pages of the psalter to early observers Their surfaces are covered with numerous horny papillae which serve to comminute the food pressed between them The abomasum is lined by a true glandular epithelium with characteristic cardiac gastric and pyloric glands

The stomach of the camel is not typical of other ruminants in that the omasum and abomasum are poorly delimited externally although internally they are quite distinct (Fig 12) Another peculiarity is the presence of sacculated glandular areas in the rumen and reticulum These have long been called water cells Fable and legend relate that the camel s ability to go for long periods without drinking is due to storage of water in these specialized structures Careful studies (A Hansen and K Schmidt Nielsen 1957) do not support this idea it is more likely that the camel s capacity to withstand drought is related to physiological mechanisms for conserving body water rather than a special arrangement for storing it

Carnivores Carnivores have a stomach with general features typical of the majority of mammals including man (Fig 11a) The region of the stomach adjacent to the esophageal opening is known as the cardia and a portion of the smooth muscle at this junction forms the relatively weak cardiac sphincter The expansible muscular fundus and body curve downward and to the right gradually narrowing to form the pyloric antrum and the somewhat more muscular pyloric canal The distal part of the encircling muscle constitutes the pyloric sphincter which controls the passage of material from stomach to duodenum The convex left side of the stomach is the greater curvature and its concave right side is the lesser curvature Although the external surface of the stomach is generally regular in contour its lining is thrown into numerous coarse predominantly longitudinal folds or rugae

Four regions of the mucous membrane generally can be recognized grossly and microscopically although their extent varies greatly in different species (Fig 12) the extension of the pale esophageal epithelium into the stomach the abrupt transition to the mucus-secreting cardiac gland region the thicker reddish brown fundic or gastric gland region which produces acid and enzymes and the mucus producing pyloric gland region

Histology The gastric wall consists of the four layers generally characteristic of the digestive tract mucosa submucosa muscularis and serosa

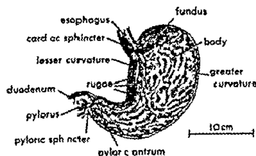


Fig 11 Anatomy of the human stomach Longitud nal sect on showing gross appearance of nterior (From L L Langley E Cherask n and R Sleeper Dynam c Anatomy and Phys ology McGraw-H ll 1958)

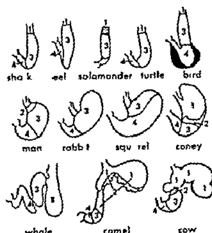


Fig 12 Diagrams to show stomach form and nature of internal l n ng in var ous mammals 1 esophageal ep thel um c l ated in salamander 2 card ac gland area found only in some mammals 3 gastric gland area 4 pyloric gland area in bird thick g zzard wall is indicated *(After Pernkopf in A S Romer The Vertebrate Body Saunders 1949)*

Although stratified squamous epithelium of the esophagus may extend for a variable distance the true gastric epithelium is a single layer of columnar cells The gastric mucin which they secrete lubricates the surface and protects it from attack by the digestive enzymes of the gastric juice In addition to rugae the lining of the stomach shows numerous fine depressions known as gastric pits (Fig 11b) into the bases of which open the numerous glands of the gastric mucosa These comparatively simple tubular structures located in the lamina propria are of three types cardiac gastric and pyloric

The cardiac glands are branched and coiled tubules composed of mucus producing cells The area which they occupy may be small (man) or quite extensive (pig)

The gastric (fundic) glands proper (Fig 11*b*) are most numerous and produce hydrochloric acid and digestive enzymes In lower vertebrates this appears to be accomplished by a single type of cell, but in mammals two major types are present chief cells for the production of enzymes and parietal cells for the production of acid Enzymes (mainly pepsin) accumulate as zymogen granules within the chief cells are discharged in an inactive state and become active when mixed with the acid gastric contents Although the gastric juice contains 0 5% hydrochloric acid it is believed that its precursor is a neutral substance produced by parietal cells and that acidification occurs by an ion exchange process as the secretion passes into the stomach The pyloric glands are somewhat simpler than the cardiac glands but like them produce only a mucous secretion

The loose and mobile submucosa is similar to that elsewhere in the digestive system The muscularis is generally arranged in three interconnected and not always well defined layers inner circular middle oblique and outer longitudinal A serosa covers the external surface of the stomach

Innervation The stomach is innervated by parasympathetic fibers of the vagus nerves and by sympathetic nerves from the celiac plexus The left vagus is distributed to the anterior wall of the stomach and the right to the posterior Impulses relayed by nerve cells in the submucous and myenteric plexuses are important in regulating acid and enzyme secretion motor activity and sphincter tone

Blood supply The stomach is principally supplied with blood by gastric branches of the celiac artery Arterial loops parallel the greater and lesser ...

the

the

Venous blood like that from the intestine returns by way of the portal vein to the liver Anastomoses occur between portal and systemic venous systems where the esophagus joins the stomach

Intestine The portion of the digestive tract beyond the stomach is the intestine It is the region of the gut wherein digestion is completed through the combined action of bile pancreatic juice and secretions from the intestinal wall itself and wherein the products of digestion are finally ab

adequate to meet the nutritive requirements of body mass the ratio of the intestine length to body length is greater in large animals than in small animals In addition plant food appears to be ... herbivores

of

es

tine about 30 ft The corresponding dimensions for man are about 22 ft and 5 ft

Small intestine In elasmobranchs and some primitive fishes the mucous membrane of the rather short small intestine is thrown into a twisted fold the spiral valve which effectively increases surface area for absorption and lengthens the path which food must travel (Fig 13) Many fishes have pyloric ceca pouchlike diverticula of the first part of the small intestine into which food may enter for digestion and absorption (Fig 6*f*) These are usually quite numerous more than 200 being found in the mackerel

Among the higher vertebrates especially mammals the mucous membrane of the small intestine is thrown into numerous thin transverse folds the plicae circulares which greatly increase the internal surface of the intestine (Fig 14*a*)

Villi The villi fingerlike outgrowths of the mucous membrane are typical of the entire small intestine in birds and mammals (Fig 14*b*) They are present in great numbers and give the surface a soft velvety appearance In man there are approximately 5 000 000 villi which provide an absorptive surface area of about 10 m² Each villus is clothed with simple columnar epithelium and has a core of delicate fibrous tissue Within each villus is a network of blood capillaries a centrally located

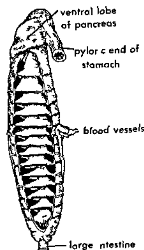


Fig 13 Small intestine of shark (*Squalus acanthias*) cut open to show spiral valve (*From C K Weichert Anatomy of the Chordates 2d ed McGraw Hill 1958*)

lymphatic capillary or lacteal and a few smooth muscle fibers The muscle fibers provide for movements of the villi which suggest a pumping action and are believed to aid the forward movement of absorbed materials which have entered the lacteals

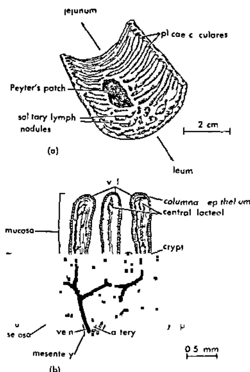


Fig 14 *Structure of the small intestine (a) Gross appearance of inner surface (b) Microscopic structure of wall shown diagrammatically (From L L Langley E Cheraskin and R Sleeper Dynamic Anatomy and Physiology McGraw Hill 1958)*

The epithelium which covers the villi and lines the small intestine is composed of three types of cells columnar absorbing cells goblet cells and argentaffin cells (Fig 15) The columnar cells are most numerous and are concerned with absorption of the products of digestion Their free surface has a finely striated appearance which electron microscopy has revealed to consist of innumerable closely packed extensions of the cell surface termed microvilli (Fig 16) Their combined surface tremendously increases the total available absorptive area Goblet cells so named because of their shape are interspersed among the columnar cells and in cyclic fashion produce and discharge mucus which serves to protect and lubricate the epithelial surface They become increasingly numerous toward the lower end of the intestine The argentaffin cells small and widely scattered in the epithelium can be recognized by their content of granules stainable with silver nitrate The function of these cells is not understood

At the bases of the villi the intestinal epithelium becomes invaginated into the lamina propria as the simple tubular intestinal glands or crypts of Lieberkuhn In addition to the types of cells already described clusters of Paneth cells (Fig 15) occur in the basal parts of the crypts These cells contain numerous coarse granules and are believed to be a source of digestive enzymes Epithelial cells are shed in large numbers from the villi and are replaced by division and upward migration of cells from the crypts of Lieberkuhn

The small intestine of mammals consists of three segments duodenum jejunum and ileum Transitions are gradual and boundaries rather arbitrary The duodenum in man is about 10 in long lies *retroperitoneally* and makes a C-shaped bend around the head of the pancreas Piercing the wall of the duodenum about 4 in beyond the pylorus are the pancreatic and bile ducts Their common site of opening is marked by a slight elevation the papilla of Vater Discharge of bile and pancreatic juice is controlled by the sphincter of Oddi In addition to villi and crypts of Lieberkuhn the duodenum of mammals is characterized by the presence of the mucus secreting glands of Brunner These lie in the submucosa and discharge their secretion through ducts which open into the bases of the crypts of Lieberkuhn

The jejunum in the average man is about 9 ft long the ileum about 13 ft Both are suspended from the posterior abdominal wall by a mesentery and together their loops fill much of the abdominal cavity (Fig 4) General structural features of the intestinal wall including villi and crypts of Lieberkuhn are similar throughout these two segments Plicae circulares reach their greatest development in the first part of the jejunum and gradually disappear below the middle of the ileum Scattered at irregular intervals beneath the epithelium of the small intestine are small solitary lymph modules In the ileum these become increasingly numerous and clumped together to form the so called Peyer's patches (Fig 14*a*) In old age Peyer's patches and lymphatic tissue in general undergo gradual involution

Blood supply The blood supply of the small intestine is derived principally from the superior mesenteric artery with a small branch from the hepatic artery contributing to the duodenum Blood returning from the intestine with absorbed food substances is transmitted by the portal vein to the liver where some of the absorbed materials are removed for storage

The parasympathetic innervation of the small intestine is by way of the vagus nerves and their connections with cells in the myenteric and submucous plexuses The vagus nerves are of primary importance in regulating intestinal secretion motility and sphincter action Sympathetic nerves from the celiac (solar) plexus and the superior and inferior mesenteric plexuses reach the intestine in close association with its blood vessels The sympathetic innervation appears to be concerned chiefly with control of blood supply and upon stimulation may also have an inhibitory influence on gastrointestinal motility

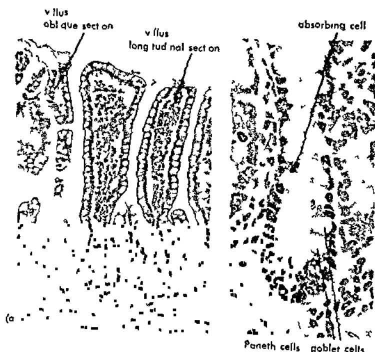


Fig 15 (a) Low power photomicrograph of a section of wall of human small intestine showing villi and crypts of Lieberkuhn (b) High power photomicrograph of crypt of Lieberkuhn (From A W Ham Histology 3d ed, Lippincott, 1957)

Large intestine The large intestine in infra mammalian species is relatively short and often not clearly delimited from the small intestine Its expanded lower end forms the cloaca a passage which is shared with the reproductive and urinary systems (Fig 7) In reptiles and birds the cloaca absorbs substantial amounts of water and the digestive and urinary waste products are discharged in a pastelike state Most reptiles and birds have one or two colic ceca protruding as blind diverticula from the large intestine (Figs 7c) These vary greatly in size and shape and are presumed to have a digestive and absorptive function

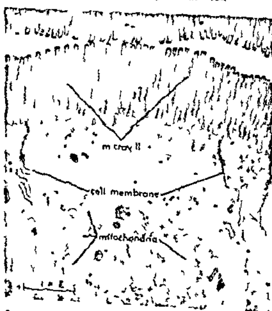


Fig 16 Microvilli on free surface of an absorbing cell from small intestine of a rat (Electron micrograph by Dr Michael L Watson)

The large intestine in mammals comprises four principal regions cecum appendix colon and rectum At the opening of the small intestine into the side of the cecum regurgitation is prevented by the ileocecal sphincter and a valvelike fold of the mucous membrane (Fig 17) The length and configuration of the cecum varies greatly in different species Herbivorous animals generally possess a relatively long and sometimes coiled cecum within which the digestion of cellulose occurs by bacterial action The human cecum lies in the right lower quadrant of the abdomen and is a relatively sho [illegible] e slender [illegible] sa of the [illegible] hatic tissue which leaves only a narrow lumen (Fig 18) Infection in this organ may progress rapidly and necessitate prompt surgical removal

The colon in man is about 5 ft long and consists of ascending transverse descending and sigmoid segments (Fig 4) The hepatic and splenic flexures which define the extent of the transverse colon lie beneath the liver and spleen respectively and are held in place by fibrous attachment to the posterior abdominal wall The outer layer of smooth muscle of the colon shows three conspicuous longitudinal bands the taeniae coli, which begin at the base of the appendix and run the full length of the colon The relative shortness of the taeniae throws the wall of the colon into a series of saccular bulges or haustra (Fig 17) Viewed from within the indentations between adjacent haustra appear as crescentic ridges termed plicae semilunares The serosa of the colon is elevated in numerous places by accumulations of fatty tissue to form irregular protrusions known as appendices epiploicae

The mucosa of the large intestine in mammals lacks villi The crypts of Lieberkuhn are longer and

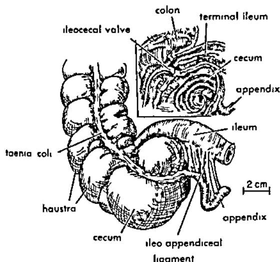


Fig 17 Junction of ileum with large intestine in man (*From L L Langley, E Cheraskin and R Sleeper, Dynamic Anatomy and Physiology, McGraw Hill, 1958*)

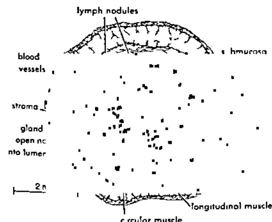


Fig 18 Cross section of human vermiform appendix (*From W M Copenhaver and D D Johnson, Bailey s Textbook of Histology, 14th ed, Williams and Wilkins 1958*)

more closely spaced than in the small intestine and goblet cells are abundant (Fig 19)

The two principal functions of the large intestine are concentration of the feces by absorption of water, and production of mucus to facilitate the forwarding of the feces

The rectum, terminal segment of the large intestine serves to hold the feces before discharge to the outside In man it is about 7 in long and can be distended to a diameter of about 3 in Its outer layer of longitudinal smooth muscle is complete and represents the expanded continuation of the taeniae coli of the sigmoid colon At its lower end the circular smooth muscle forms an internal (involuntary) sphincter (Fig 20) Below this is the external (voluntary) sphincter of striated muscle Bands of circular smooth muscle near the middle of the rectum throw the mucosa into two or three transverse ridges, the plicae transversales recti which help to support the rectal contents and relieve pressure on the anal sphincters The mucosa of the lower part of the rectum has a number of longitudinal folds, the rectal columns, of which the lower ends are united just above the anal orifice by smaller transverse folds forming the anal valves At this juncture the columnar epithelium lining the rectum abruptly gives way to the stratified squamous epithelium lining the lower anal canal and becomes continuous with the epidermis of the external skin

Blood supply The blood supply of the large intestine is derived from the superior and inferior mesenteric arteries and branches of the internal iliac arteries to the rectum Returning blood goes by way of the portal system to the liver Veins in

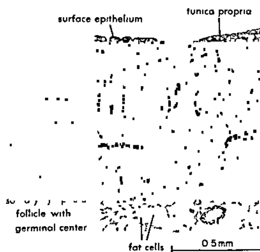


Fig 19 Transverse section through mucosa and submucosa of human colon (*After Braus in W M Copenhaver and D D Johnson, Bailey s Textbook of Histology, 14th ed, Williams and Wilkins, 1958*)

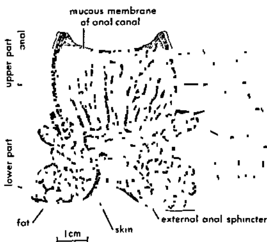


Fig 20 Mucous surface of anal canal in man (longitudinal section) (*From W M Copenhaver and D D Johnson, Bailey's Textbook of Histology, 14th ed liams and Wilkins, 1958*)

the submucosa of the anal canal have a propensity for becoming dilated and painful a condition known as hemorrhoids

Innervation Parasympathetic innervation of the large intestine is supplied jointly by the vagus nerves and the sacral division of the parasympathetic system Sympathetic innervation is from the superior and inferior mesenteric and hypogastric plexuses Emptying of the rectum is largely a reflex act regulated by the parasympathetic nerves Innervation of the external sphincter by the voluntary nervous system places evacuation of the rectum under voluntary control [V M E]

PHYSIOLOGY

The foodstuffs of nature with which animals satisfy their energy requirements exist as carbohydrates proteins and fats These must be digested respectively to their component sugars amino acids and fatty acids in order to be utilized by the body tissues Digestion among the vertebrates is entirely extracellular and is carried out within the lumen of the alimentary tract The alimentary tract consists of a tube which passes the length of the trunk and is characteristically coiled Glands (pancreas liver and less commonly the salivary) attached to the tract by means of ducts supply food softening fluids and digestive enzymes for chemically splitting the molecules of foodstuffs *See* DIGESTIVE GLAND, ORAL GLAND

The anterior end of the alimentary canal is equipped in the cyclostomes with a sucking device for obtaining fluid food and in the gnathostomes with jaws and teeth or a bill for grasping or comminution of food The canal has a muscular coat which triturates the food mixes it with digestive enzymes and provides peristaltic propulsion of the food mass The digestive enzymes are physiologically similar throughout the vertebrates but slight variations in chemical properties between enzymes from different species have been noted The amounts of each enzyme produced vary with the species and are functionally adapted even to subtle differences in eating habits For example the proteases are abundant in carnivores, whereas the herbivores are well supplied with carbohydrases

Mouth. The primary function of the mouth in digestion is the reception and swallowing of food Other important functions subserved by it in some but not all vertebrates are the mastication and moistening of whole foods Actual digestion in the mouth is of limited occurrence The functions of the mouth in the digestive process, then, are mainly mechanical, that is, prehension, tearing, cutting, and grinding of food and mixing it with a fluid medium

Salivary gland secretions The fluids emptied into the oral cavity are produced by the salivary glands widely present in vertebrates Although they generally secrete a lubricating medium for the food mass, in certain lower forms they are specialized for the production of toxins or anticoagulants important to the gathering of food The salivary glands may secrete either a mucous or a serous fluid or a mixture of these The sole digestive enzyme produced by salivary glands is salivary amylase (ptyalin) This has been found in significant quantity only in birds, man, apes, elephants and pigs Its presence in rodents, dogs, ungulates and frogs is doubtful or slight Man has three pairs of discrete salivary glands the submaxillary, sublingual and parotid



Fig 21 The salivary glands (J P Schaeffer, ed, Morris' Human Anatomy, 11th ed, Blakiston, 1953)

Salivary amylase Salivary amylase appears to be an α amylase in common with the amylases found in pancreatic juice, blood, urine, and somewhat rarely in bile The polysaccharides plant starch and animal glycogen, are split by this enzyme Amylase action is purely hydrolytic and entails a cleavage of the α glucosidic 1,4 linkage The degradation products are always a fermentable sugar, chiefly maltose, and a nonfermentable dextrin One molecule of starch may give rise to as many as 300–500 molecules of glucose Salivary amylases from different species vary considerably in their pH optima, that for man is 6 2 Salivary amylase like other animal amylases, is activated by chloride ions and loses its activity on dialysis Salivary digestion is inhibited by acids, and hence ceases when the food reaches the stomach and is exposed to the highly acid gastric juice Salivary amylase, although very rapid in its action is not of critical importance to the total economy of any of the animals in which it is found Its absence does not result in an impairment of digestion *See* AMYLASE, CARBOHYDRATE METABOLISM

Saliva Saliva is secreted in much greater volume than is generally appreciated Man produces 0 5–1 liter per day but horses and cows subsisting on dry forage may produce 40 and 60 liters per day respectively Climactically, however, a single parotid gland of sheep has been shown to yield 1–1¾ liters of fluid per hour Much of the water supplied during alimentation is absorbed during the formation

of feces The flow of saliva is controlled entirely by the nervous system Innervation of the salivary glands is supplied both by the sympathetic and the parasympathetic nervous systems The flow of saliva is sustained at basal levels wholly involuntarily by sensations arising in the mouth The sight smell or even the thought of food will also excite a copious flow of saliva As a result of the classic experiments carried out by I Pavlov it has been well established that the body has an elaborate mechanism for increasing the flow of saliva in association with foods and eating Pavlov offered food in precise temporal relationships to various external stimuli during conditioning periods and was able to demonstrate an increased salivary flow in response to such an external stimulus as the ringing of a bell

Composition of saliva Saliva contains 97 99 5% water other components are mucin buffering salts amylase in the instances noted above and less constantly certain other substances such as antibacterial principles including lysozyme and a mildly hemostatic agent The last two components account perhaps for the wound healing properties which ancient lore has ascribed to saliva

The large input of this watery fluid at the entrance to the digestive tract serves several valuable functions By solvent action it makes the tasting of dry foods possible and thus stimulates increased salivary flow Saliva also provides the requisite aqueous medium for bringing the several digestive enzymes in contact with fine food particles It speeds the clearing of food particles from the mouth, softens the food and facilitates its movement along the digestive canal The viscosity of saliva is due to its mucin content The demulcent effect of mucin keeps the buccal membranes free from friction between opposing surfaces and from friction with food during repeated swallowing It also provides protection to the lining of the esophagus against the friction caused by hard food particles The value of saliva as a wetting and lubricating agent is commonplace until its absence as in febrile states is painfully apparent

Teeth Aside from their role as biologic weapons of defense the teeth serve primarily the functions of prehension cutting and mastication of food The incisors and canines with their sharp incisal or conical surfaces are well adapted for prehension and cutting, whereas the molars with their broad occlusal surfaces are admirably suited for the grinding action of mastication Beyond question mastication is an aid to digestion simply on the basis that it increases the surface area of food exposed to digestive action However data for properly evaluating this function are not available Although there can be no doubt that food swallowed whole as by snakes and erratically by other forms is subject to complete digestion the wisdom of nature would dictate that a complex dentition as in the mammals is not just a convenience but an integral part of the whole organism The evidence for man although inadequately documented favors overwhelmingly the view that mastication is a meaningful function and that the teeth as well as the salivary glands play an important role in the long term maintenance of optimum health The incompleteness of the knowledge of the physiology of teeth and saliva in man is a measure not of their utility to the basic organismal economy but of the neglect of oral structure as subject for physiologic investigation The new standards of human health have sharpened the urgency for extensive study of these difficult problems *See* DENTITION TOOTH [R O C]

Tongue The tongue is composed principally of skeletal muscle It is enveloped by mucous membrane except at the root through which the extrinsic muscles blood vessels and nerves enter The muscle of the tongue itself is innervated by the hypoglossal or XIIth cranial nerve The mucous membrane on the undersurface forms a fold termed the frenulum linguae or more commonly simply the frenulum which binds the tongue with the exception of the tip to the floor of the mouth In some individuals the frenulum is so short that normal movements of the tongue are unduly restricted and thus a speech impediment results This condition is technically known as ankyloglossia and popularly as tongue-tie In some instances movement of the tongue is so limited that the infant is unable to suckle *See* SPEECH

Normally the upper surface of the tongue presents a moist pink appearance Under other conditions it may appear yellow and dry because of the accumulation of shed epithelial cells remains of food and organisms The surface of the tongue is characterized by the presence of papillae

The taste buds are scattered over the upper surface of the tongue Those of the anterior two-thirds of the tongue are innervated by the facial or VIIth cranial nerve whereas the remainder receive fibers from the glossopharyngeal or IXth cranial nerve *See* TONGUE

Pharynx The pharynx is a tube about 5 in long lined with mucous membrane which communicates with the nose and mouth above and the larynx and esophagus below The pharynx is separated from the mouth by the palatoglossal and palatopharyngeal arches called the fauces Between these arches are located the palatine tonsils *See* PHARYNX TONSIL

Esophagus The esophagus or as it is sometimes called the gullet is the most muscular part of the entire alimentary canal It extends from the termination of the pharynx to the cardiac orifice of the stomach The esophagus varies in length from 10 to 12 in and in breadth from ½ to 1 in or more The mucous membrane lining is composed of stratified squamous epithelium

In the neck the esophagus is situated just in back of the trachea and is connected by loose tissue to the trachea Behind the esophagus is the vertebral column In the thorax the esophagus is

crossed by the left bronchus and descends in association with the descending thoracic aorta The esophagus ultimately passes through the esophageal orifice in the diaphragm to which it is connected It then merges with the stomach *See* ESOPHAGUS

Gustation The taste buds located mostly on the surface of the tongue are chemoreceptors and as such are attuned to the chemical composition of the oral cavity There are specific substances which will stimulate only one of the four types of taste buds It is thus possible to map the taste receptors

It is common experience that more than four basic tastes can be easily differentiated This ability depends upon two factors utilization of more than one type of taste bud in a blend so to speak and derivation of clues from activation of extragustatory receptors such as those for temperature pressure touch and olfaction For example if an individual were to chew a piece of uncooked potato he could distinguish it from a raw carrot or an apple Yet it is clear that none of these foods can be described as truly acid or sour bitter or sweet The texture of the substance its moisture content the temperature the relative proportion of the various types of taste buds stimulated and finally the sense of smell all contribute This information is then transmitted to the sensory cortex and in the light of past experience a mental image results *See* TASTE

It is easy to demonstrate that the taste of a substance depends on contrast For example coffee consumed after a very sweet dessert does not taste as sweet as the same beverage does in the morning when nothing else has preceded it

Deglutition The act of swallowing is termed *deglutition* It is a complex process involving both voluntary and reflex mechanisms In order to analyze this act it is best to consider it as occurring in three stages

In the first stage of deglutition the food is ground and rolled into a mass called a bolus and thoroughly soaked with saliva The tongue then directs the bolus to the back of the mouth and forces it to enter the pharynx All these acts are purely voluntary the individual can chew the food as long or as briefly as he desires Once he forces the bolus into the pharynx however the voluntary phase of swallowing has ended and all subsequent mechanisms are purely reflex The reflex control of the second and third stages is initiated by the voluntary act of the first stage

During the second stage of deglutition the bolus passes through the pharynx to enter the esophagus The passage of food through the pharynx is in reality rather precarious There are many orifices into and out of the throat the opening of the trachea through which air but not food must pass the entrances of the eustachian tubes which connect the middle ear and the pharynx the openings into the mouth and nasal passageways and finally the esophageal opening

The bolus of food must be directed into the ... none of the others In ... cular ... leaving the esophagus free to receive the bolus The bolus is prevented from reentering the mouth by the tongue and neighboring structures The soft palate is raised to bar it from the nasal passages The orifices to the eustachian tubes are guarded by muscles which prevent entrance The trachea is protected by elevation of the larynx and movements of the epiglottis All these actions must take place within a split second and synchronously with the contractions which are responsible for propelling the food through the pharynx

In the third phase of deglutition the food traverses the esophagus to enter the stomach Man usually eats in the upright position, therefore the force of gravity assists in transporting the bolus through the esophagus However the force of gravity is not indispensable to the swallowing act It is possible for an individual to swallow and pass the bolus or even liquids through the esophagus while in an upside down position

The esophagus is composed chiefly of muscles which contract in wavelike fashion along the length of the tube These are called peristaltic waves or movements The term peristalsis means clasping and compressing and describes the contraction of one part of the tube then contraction below it and relaxation of the originally constricted segment and so forth (Fig 22) Thus the bolus is moved through the esophagus to enter the stomach [L L L]

Stomach The digestive processes are similar in all vertebrates this is particularly true of the basic stomach functions The vertebrate in which the physiology of the stomach has been most extensively studied is the dog and studies on man rat

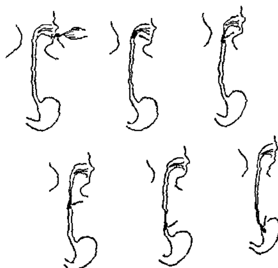


Fig 22 Esophageal peristalsis Stages in passage of bolus from the mouth to the stomach (From L L Langley E Cheraskin and R Sleeper Dynamic Anatomy and Physiology McGraw Hill 1958)

and other animals indicate that the dog is typical of most mammals in this respect The higher vertebrates whose stomachs show the greatest differences from the simple stomach of the dog are the birds and the ruminants This article is largely concerned with the ruminants which are of great economic importance, less information is available about stomach modifications in lower vertebrates

Despite its important role in digestion the stomach is not essential to life Gastrectomized animals are subject to certain extra stresses but a high percentage survive

Functions The stomach is a bag shaped organ into which the food passes from the esophagus The partially digested food is afterwards released piecemeal to be dealt with by the digestive processes in the small intestine The stomach comprises the corpus and a smaller more posterior portion called the antrum These regions are also known as the cardiac and pyloric portions Glands in the stomach wall secrete mucus partly to protect the stomach wall but also to moisten the food

The main function of the stomach is to initiate the digestion of protein This digestion is only partial but is sufficient to disrupt the cells of the food and release the intracellular contents rendering them susceptible to digestion in the small intestine The gastric juice responsible for protein digestion is secreted by the fundic glands and comprises hydrochloric acid which is secreted by the parietal cells and is at pH 0 9–1 0 and pepsinogen which is secreted by the chief cells The hydrochloric acid activates this material to yield pepsin the active form of the enzyme At the low pH produced by the hydrochloric acid pepsin is a potent proteolytic enzyme

Carbohydrate digestion is negligible It is possible however that some sucrose is hydrolyzed by the hydrochloric acid

Fat digestion proceeds to some extent The stomach secretes a small amount of lipase which is adequate to work on finely divided fat as found for example in egg yolk Some more active lipase is frequently regurgitated from the duodenum

Precipitation of casein from milk is accomplished by the enzyme rennin which is found in significant concentrations only in the gastric secretion of infant animals Pepsin and hydrochloric acid will also precipitate casein The mixing and trituration of food is carried on in the stomach by powerful muscular contractions

Absorption of water salts alcohol and glucose proceeds to a very limited extent in the stomach These materials are usually in liquid form and pass through the stomach to the duodenum so rapidly that their total absorption is negligible

Gastric secretion is under nervous and hormonal control and has been studied extensively It is conveniently divided into three phases, the psychic gastric and intestinal During the psychic phase higher centers are stimulated by the sight of food or even by associations of feeding Nervous impulses then initiate secretion The gastric phase is induced by the mechanical stimulation of the mucosa by food and by the liberation of hormones from the mucosa into the blood stream In the intestinal phase a hormone is secreted into the blood by the intestinal wall after food reaches the small intestine and this induces further gastric secretion The above functions are seen typically in the vertebrate stomach Exceptions and extensions to these are discussed below *See* PROTEIN METABOLISM

Lower vertebrates The cyclostomes (lampreys and hagfish) have no stomachs and in many of the fish and amphibians the stomach is merely a local widening of the gut In carnivorous species, however the stomach is important for storage This is because the prey is frequently swallowed whole and must remain in the stomach for a long time for gastric digestion For example, whole fish remain in the pike s stomach for as long as 4 days In these species the hydrochloric acid serves to kill the prey and to prevent putrefaction in addition to its role in protein digestion The acid secreted by fish is nevertheless slightly less strong than in mammals so that the gastric juice undiluted by food has a pH of about 2 0 as compared with 0 9 in the dog In some fish the flow of gastric juice is apparently constant but in most it is thought to be stimulated by the mechanical contact of food with the gastric mucosa No nervous control of gastric secretion has been demonstrated in fish or amphibians although injections of human gastrin will promote secretion in the frog

Reptiles Among the reptiles the processes are similar to those described for the fishes except that *Crocodilia* have a primitive gizzard The anterior portion of the stomach is lined with a horny material and is immensely muscular With the assistance of stones swallowed by the animal powerful contractions of the stomach muscles grind the food The ground up material then passes through an aperture into the smaller posterior part of the stomach which contains the digestive glands

Birds The characteristic modification of the stomach in birds is also a gizzard but in this case it comprises the posterior portion of the stomach The anterior portion is the proventriculus which is glandular and similar in function to the dog's stomach The similarity extends even to the control of gastric secretion gastric flow in ducks has been conditioned to the ringing of a bell in the same fashion as in the experiments of Pavlov with the dog In most birds there is probably a minimum of digestion in the proventriculus because the food reaches it in an undivided state and it is too small for prolonged storage The food is rapidly passed on to the gizzard where it is triturated for 1–12 minutes depending on the consistency of the food This mixes the food thoroughly with the digestive enzymes previously secreted in the proventriculus, digestion continues in the duodenum In birds the biliary and pancreatic ducts enter the duodenum at its distal end so that the full length of the duodenum is available for gastric dige

The gizzard is most highly developed in seed eating birds least so in birds of prey The herring gull shows seasonal variation in the hardness of the gizzard associated with dietary changes from grain in the summer to fish in the winter Some experiments have shown that the presence of grit in the gizzards of domestic fowl is less important than might be supposed Nevertheless birds do indeed consume grit and small rocks which are retained in the gizzard The gizzard is of advantage to birds because it enables them to utilize the high nutrient value of plant seed Mammals do not derive the full benefit from this food source unless the seeds are first ground (as for livestock) or cooked

Mammals and marsupials In mammals the stomach functions are typically as described for the dog In the primates pigs and some rodents ptyalin is present in the saliva and the digestion of starch begun in the mouth is continued to some extent by this enzyme after reaching the stomach It is stopped by the low pH of the gastric secretions but after a large meal the hydrochloric acid may take 30 minutes to penetrate the whole of the food mass within the stomach Only in man is the ptyalin concentration of the saliva great enough for this to be an important contribution

In man the psychic phase of gastric secretion is modified by factors more complex than the mere *imminence* of a meal Gastric secretion is determined to some extent by the emotional state of the individual and there is evidence that hypersecretion as a result of emotional stress is involved in the development of duodenal ulcers In older people the absence of stomach acid and pepsin has a high correlation with the incidence of gastric carcinoma

It is of interest to note that the herbivores and rodents do not vomit

The beaver wombat and koala all possess a cardiac gland situated on the lesser curvature of the stomach In the case of the beaver it has been proposed that this gland secretes an enzyme capable of digesting the cellulose in the tree bark which is the principal food of this animal However no such digestion has been demonstrated experimentally

Some of the pelagic mammals such as whales and sea cows have the pyloric end of the stomach modified to form a long cecumlike structure The physiology of this is obscure Seals' stomachs are often found to contain pebbles but it is not known whether these serve a digestive function

Ruminants Of all the vertebrates the ruminants show the most spectacular modification of the stomach The anterior portion is enlarged to form a fermentation vat in which microorganisms break down the otherwise indigestible cellulose of the animal s vegetable diet and make the products of the fermentation available for the nutrition of the host animal *See* METABOLISM IN RUMINANTS

The ruminants include cows (and wild relatives) sheep goats deer musk oxen camels and llamas The chevrotains (Malayan mouse deer and others) are small animals in which the rumen development is less complete

In the cow the stomach has four compartments (see Fig 23) The anterior two are the rumen and reticulum which together form the fermentation vat Their total volume may be 80 gallons in a large cow The pH in the rumen is about 6 0–7 5 and is partly buffered by the bicarbonate content of the saliva which is continually swallowed in copious amounts The rumen has a relatively smooth lining but the reticulum as the name implies has an inner surface resembling a honeycomb In the camel the cells of the reticulum are particularly deep The third compartment of the stomach is the omasum The walls of this much smaller compartment are muscular and thrown into leaflike folds which practically fill the cavity Fermented material passing through the omasum is subjected to squeezing and trituration on passage The omasum discharges into the abomasum which alone contains digestive glands and is generally similar to the simple stomach of other mammals The rumen reticulum and omasum are traditionally referred to by husbandmen as the paunch honeycomb and manyplies respectively

When grazing ruminants swallow the grass rapidly and it remains in the rumen for some time Later boli of rumen contents are regurgitated and chewed more thoroughly before being returned to

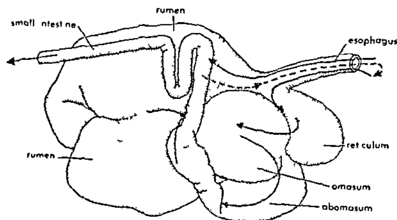


Fig 23 Stomach of the cow (From T A Rogers The metabolism of ruminants Sci American 198(2) 34–38 1958)

the stomach This permits fermentation to begin disrupting the cellulose structure before it is chewed It is also likely that this practice of rumination has survival value in the wild state where the animal can rapidly eat its fill on the open plain and then work it over in some safe inaccessible retreat

At one time it was thought that the only function of the microorganisms of the rumen reticulum was to break down cellulose into compounds which could be completely digested in the abomasum and small intestine Studies conducted since about 1940 have shown however that the activities of the symbiotic microflora are even more far reaching The results of these investigations are summarized below

Cellulose (and other polysaccharide material such as starch) is fermented to yield large quantities of short-chain fatty acids including acetic acid in the greatest quantity as well as important amounts of propionic butyric and some branched chain acids These are absorbed directly into the blood stream through the wall of the rumen and supply the bulk of the animal's energy requirements This is in contrast to the negligible absorption of nutrients in simple stomachs

The active fermentation of all polysaccharide material means that the ruminant absorbs little or no glucose from the gut This with the plentitude of acetic acid is reflected in the intermediary metabolism of the animal The basic unit of energy metabolism is acetate rather than glucose and the glucose which is required by the animal for lactose synthesis is synthesized from the propionate and butyrate absorbed from the rumen

The fermentation involves the protein content of the diet as well as the polysaccharides and some of this is degraded to its constituent ammonia and amino acid residues These simple materials are utilized by some microorganisms in the synthesis of their own structural proteins which become available to the cow when the symbiotes are themselves digested in the abomasum and small intestine The "reshuffling" of the nitrogen in the rumen results in the manufacture of protozoal protein of high biological value from relatively inferior plant protein That is the protozoal protein contains more of the amino acids which the higher animals are unable to synthesize for themselves Thus the synthetic activities of the rumen microflora make ruminants independent of a supply of these essential amino acids in their diet Modern stock feeders take advantage of the nitrogen fixing symbiotes by including a small proportion of urea in their cattle feed The microorganisms can synthesize protein from this nonprotein nitrogen See METABOLISM IN RUMINANTS

The rumen microbes similarly synthesize the water soluble vitamins for their own use from simple materials This supply is enough to meet all the requirements of the host's tissues even in the absence of these vitamins in the diet

The fermentation produces large amounts of carbon dioxide and methane and these gases must released from the rumen by belching Any interference with the belching mechanism causes the animal to bloat and die

The infant ruminant has an undeveloped rumen and during the first weeks of life a calf resembles mammals with simple stomachs in its digestive processes intermediary metabolism and dependence on a dietary supply of essential amino acids and vitamins As it begins to eat hay the rumen enlarges the microflora develops and it gradually changes over to the characteristics of the adult ruminant It is likely that oral contact with older animals helps to inoculate the appropriate organisms in the infant rumen

It is evident from this brief review that the modification of the anterior portions of the ruminant stomach is a great advantage to these animals In addition the ability of domestic ruminants to produce milk and meat from grass is the very foundation of animal husbandry and of inestimable value to mankind [T A R]

Small intestine Digestion in the small intestine is the chemical process whereby food is converted into substances which may be absorbed and assimilated by the body Satisfactory digestion in the small intestine that is in the duodenum the jejunum and the ileum depends on the secretions of the pancreas liver and intestinal glands which provide liquid for dilution and solution of the food a reducer of surface tension for the emulsification of fat and enzymes for the breakdown of foods into simple molecules and on the motor action of the small intestine which mixes the food with the secretions and propels the contents in a caudal direction

The stomach acts as a digesting reservoir for the alimentary canal It delivers its contents after their partial digestion in spurts into the duodenum By the time food reaches the duodenum its state is considerably changed from when it was eaten Some of these changes are essential for satisfactory digestion in the small bowel The food has been diluted by the salivary and gastric secretions and has been strongly acidified by the latter Before the salivary amylase is inhibited by the increasing acidity of the gastric content 60–70% of the ingested starch has been converted to maltose By the time the proteins of the food reach the duodenum they too are partially digested Hydrochloric acid and pepsin are combined in the stomach to simplify many proteins to proteoses and peptones Fat however is not altered while it is in the stomach

The mixture which is delivered to the duodenum is strongly acid with a pH of 1 5–2 One of the first jobs to be done by the duodenum is to bring the contents to or close to neutrality This is essential for the digestive processes of the small intestine which all require a nearly neutral pH for optimum activity Neutralization is accomplished by the alkaline secretions of the duodenum and pancreas aided by the bile from the liver De[...]ery of the acid gastric contents into the d[...] [...]trolled so that these neutralizing [...] be overwhelmed Acidification of

denum results in a prompt reduction or complete cessation of propulsive gastric motility so that further delivery of acid material is delayed until neutrality is restored in the duodenum

Fat introduced into the duodenum also delays gastric emptying This too is a protective mechanism for too rapid delivery of fat into the small bowel may overwhelm its capacity for digesting and absorbing lipids and lead to fatty diarrhea

Excessive concentrations of sugars in the duodenum also will delay delivery of the gastric contents into the duodenum Again a function of the first part of the small bowel is being protected The duodenum acts to bring the total concentration of dissolved substances in its contents to that of blood or nearly so When the solutions are very concentrated some time may be required delivery from the stomach is slowed so that this function will not be swamped

How these three regulatory actions are accomplished is not known exactly Certainly neural and possibly humoral factors are involved Removal of the vagal nerve supply to the stomach eliminates a part of the normal balance and leads to reduced gastric motility and retention of food and secretions in the stomach

Once neutrality and osmotic equilibration are accomplished in the duodenum digestion proceeds quickly It continues as the contents traverse the small bowel and is complete by the time they are delivered to the large bowel Absorption commences as soon as the nutrients have been sufficiently simplified to pass through the mucosal lining of the small bowel to enter the blood and lymph Digestion is accomplished by the combined actions of pancreatic juice bile and the intestinal secretions

Pancreatic secretion The pancreas a branched gland is situated along the side of the duodenum in all vertebrates Pancreatic secretions enter the duodenum by one or more ducts a short distance from the pylorus (Fig 24) See PANCREAS

Composition of secretion Pure pancreatic juice is a colorless clear or opalescent rather viscous fluid with a pH of 8 4–8 9 The inorganic constituents include sodium potassium and bicarbonate ions the bicarbonate renders the juice alkaline The organic constituents consist of various proteins These can be precipitated from fresh juice by heat or alcohol The proteins contain the enzymes of the juice The three main enzyme systems in pancreatic secretions are proteolytic amylolytic and lipolytic

Action on food Trypsin is the main proteolytic enzyme formed by the pancreas It is secreted in an inactive form called trypsinogen Trypsinogen does not become active until it enters the duodenum This is a safety mechanism to prevent autodigestion of the pancreas The secretions of the intestinal glands always contain an enzyme called enterokinase which splits a polypeptide from the trypsinogen at pH 5 0 to form active trypsin Trypsin itself also can activate trypsinogen Trypsin acts on proteins some of which have been denatured by the gastric acid to form proteoses peptones polypeptides and amino acids Trypsin is in fact not a single enzyme but a group of enzymes each member of which attacks the peptide molecule in a specific manner Some polypeptides resist digestion by trypsin even if they remain with it for a long time

Chymotrypsinogen is another proteolytic enzyme that is secreted in inactive form in the pancreatic juice It is activated by trypsin to form chymotrypsin This enzyme clots milk and hydrolyzes casein and gelatin

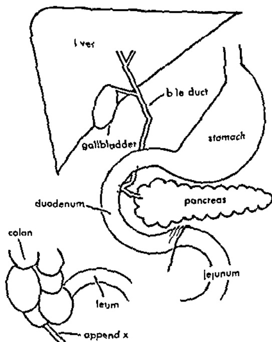


Fig 24 Small intestine in digestion (Mayo Clinic Rochester Minn)

Carbohydrates are acted upon by pancreatic amylase or amylopsin which acts like salivary amylase to hydrolyze starch first to dextrin and then to maltose Amylopsin has maximal activity at pH 7 0 A maltase capable of splitting maltose into glucose has been identified in pure pancreatic juice

Fats are subjected to the action of pancreatic lipase or steapsin which hydrolyzes neutral fat into glycerol and fatty acids its maximal activity is at pH 8 0 This digestion of fat is considerably enhanced by the concomitant action of bile salts on the duodenal contents Bile salts lower the surface tension between water and neutral fat so that the fat can become emulsified into minute globules Thus the surface area of fat available for enzyme activity is increased greatly and the rate of hydrolysis is accelerated The fat splitting power of pancreatic juice is trebled by the presence of bile salts Fat also aids its own digestion Liberated fatty acids unite with alkali in the bowel to form sodium salts or soaps that have the same emulsifying action as the bile salts

Regulation of secretion In the intact animal nervous and hormonal mechanisms combine to

produce the pancreatic juice with the proper composition for the most effective digestion of foodstuffs in the small bowel Food in the mouth acts as a stimulus for the reflex production of pancreatic juice This reflex is of vagal origin and the juice is rich in enzymes

In a classic experiment in 1902 W Bayliss and E. Starling discovered that placing acid in an isolated denervated loop of jejunum resulted in the secretion of pancreatic juice An acid extract of the jejunal mucosa was neutralized and injected intravenously and again the pancreas began to secrete They called the active agent in their extract secretin Acid gastric juice in the duodenum is the physiologic stimulus for release of secretin Secretin is a polypeptide of low molecular weight which is destroyed by trypsin and pepsin Secretin has been identified in birds mammals and some nonmammalian vertebrates It stimulates the cells of the pancreatic glands to produce a juice which is strongly alkaline and watery and is low in enzymes By stimulating the production of such a juice secretin aids in the neutralization of the acid that led to its production Secretin may further reduce acidity by direct action on the gastric glands Intravenous injections of secretin inhibit the production of acid by the stomach

Neutralization is essential because the enzymes of the pancreatic juice require a near neutral pH for their most effective digestive action Once the contents are neutralized food in the duodenum and upper part of the jejunum releases pancreozymin into the circulation In contrast to secretin, pancreozymin evokes a pancreatic secretion which is thick and viscid and is high in enzymes

The pancreas is innervated by vagal and sympathetic nerves Vagal stimulation produces a viscid juice which is rich in enzymes Vagal nerve impulses also potentiate the effects of secretin and pancreozymin The role of the sympathetic nerves is not well defined Because stimulation of sympathetic nerves causes constriction of pancreatic blood vessels the action of these nerves may reduce pancreatic secretion

Bile The liver secretes bile continuously It is collected in a ductal system which ends in the common bile duct This duct opens into the first part of the duodenum in the same region as the pancreatic duct so that bile and pancreatic juice come into contact with the food simultaneously shortly after its entrance into the duodenum

The majority of vertebrates possess a gallbladder for storage and concentration of bile Some animals such as the rat and horse do not and in these the bile ducts have a greater capacity The coordinated relaxation of the sphincter at the lower end of the common bile duct and the contraction of the gallbladder allow bile to enter the duodenum in response to demand *See* GALLBLADDER LIVER

Composition of secretion Fresh bile from the liver is an orange colored alkaline fluid Bile which has been stored in the gallbladder is darker and more viscid because of concentration and the addition of mucus by the gallbladder The principal components of bile are the bile salts bile pigments cholesterol and inorganic salts. The bile salts and cholesterol have digestive function while the other components are waste products for excretion The bile salts are the sodium salts of glycocholic and taurocholic acid They promote the emulsification of fat in the intestine and also keep cholesterol and lecithin in solution Ninety per cent of the bile salts are reabsorbed in the small intestine and remetabolized by the body

Cholesterol is an unsaturated secondary alcohol closely related to many hormones It is a constituent of all cells and in the small intestine aids in the emulsification and absorption of lipids It also is reabsorbed in the presence of bile *See* CHOLESTEROL LIPID METABOLISM

There are two bile pigments bilirubin which gives the yellow color to bile and biliverdin which adds the green tint Bilirubin is formed during the breakdown of hemoglobin by the reticuloendothelial system in the liver and other parts of the body Biliverdin is an oxidation product of bilirubin Bilirubin is changed in the intestine to urobilinogen part of which is reabsorbed while the rest is further changed to stercobilinogen This pigment is brown and colors the intestinal waste products *See* BILE ACID BILIRUBIN

Action on food Apart from its role in the digestion of fat whole bile in the intestine facilitates the absorption of the fat soluble vitamins D K and E and has the property of increasing intestinal motility *See* VITAMIN

Control of secretion Protein or fat in the duodenum stimulates the flow of bile which reaches a maximum in 1 hour and lasts 2–3 hours Carbohydrate has no effect on the secretion of bile Bile and bile acids in the duodenum however are potent stimulators of the further secretion of bile

The release of bile is under neural and hormonal control Stimulation of the vagal nerve releases bile by causing contraction of the gallbladder and relaxation of the sphincter at the lower end of the duct Acids fats, or egg yolk when placed in the intact intestine or in an isolated denervated loop of bowel cause contraction of a denervated gallbladder This indicates a hormonal mechanism and cholecystokinin is the name given to the responsible substance Cholecystokinin has been separated as a fraction from extracts of the mucosa of the upper part of the small intestine

Intestinal secretion Secretion of intestinal juice and absorption of products of digestion take place together throughout the whole small intestine The upper part of the duodenum contains special glands the glands of Brunner which are more numerous in herbivores than in carnivores These glands secrete an alkaline mucus that contains amylase and enterokinase One of the functions of this secretion is to protect the intestinal mucosa from erosion by gastric acid

The intestinal glands proper are present throughout the whole small intestine and consist of groups of secreting cells situated around the intestinal villi

denum results in a prompt reduction or complete cessation of propulsive gastric motility so that further delivery of acid material is delayed until neutrality is restored in the duodenum

Fat introduced into the duodenum also delays gastric emptying This too is a protective mechanism for too rapid delivery of fat into the small bowel may overwhelm its capacity for digesting and absorbing lipids and lead to fatty diarrhea

Excessive concentrations of sugars in the duodenum also will delay delivery of the gastric contents into the duodenum Again a function of the first part of the small bowel is being protected The duodenum acts to bring the total concentration of dissolved substances in its contents to that of blood or nearly so When the solutions are very concentrated some time may be required delivery from the stomach is slowed so that this function will not be swamped

How these three regulatory actions are accomplished is not known exactly Certainly neural and possibly humoral factors are involved Removal of the vagal nerve supply to the stomach eliminates a part of the normal balance and leads to reduced gastric motility and retention of food and secretions in the stomach

Once neutrality and osmotic equilibration are accomplished in the duodenum digestion proceeds quickly It continues as the contents traverse the small bowel and is complete by the time they are delivered to the large bowel Absorption commences as soon as the nutrients have been sufficiently simplified to pass through the mucosal lining of the small bowel to enter the blood and lymph Digestion is accomplished by the combined actions of pancreatic juice bile and the intestinal secretions

Pancreatic secretion The pancreas a branched gland is situated along the side of the duodenum in all vertebrates Pancreatic secretions enter the duodenum by one or more ducts a short distance from the pylorus (Fig 24) See PANCREAS

Composition of secretion Pure pancreatic juice is a colorless clear or opalescent rather viscous fluid with a pH of 8 4-8 9 The inorganic constituents include sodium potassium and bicarbonate ions the bicarbonate renders the juice alkaline The organic constituents consist of various proteins These can be precipitated from fresh juice by heat or alcohol The proteins contain the enzymes of the juice The three main enzyme systems in pancreatic secretions are proteolytic amylolytic and lipolytic

Action on food Trypsin is the main proteolytic enzyme formed by the pancreas It is secreted in an inactive form called trypsinogen Trypsinogen does not become active until it enters the duodenum This is a safety mechanism to prevent autodigestion of the pancreas The secretions of the intestinal glands always contain an enzyme called enterokinase which splits a polypeptide from the trypsinogen at pH 5 0 to form active trypsin Trypsin itself also can activate trypsinogen Trypsin acts on proteins some of which have been denatured by the gastric acid to form proteoses peptones polypeptides and amino acids Trypsin is in fact not a single enzyme but a group of enzymes each member of which attacks the peptide molecule in a specific manner Some polypeptides resist digestion by trypsin, even if they remain with it for a long time

Chymotrypsinogen is another proteolytic enzyme that is secreted in inactive form in the pancreatic juice It is activated by trypsin to form chymotrypsin This enzyme clots milk and hydrolyzes casein and gelatin

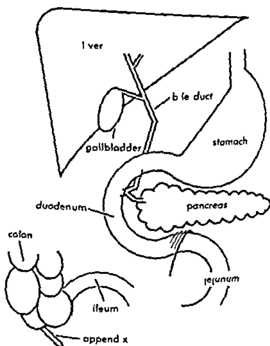


Fig 24 Small intestine in digestion (Mayo Clinic Rochester Minn)

Carbohydrates are acted upon by pancreatic amylase or amylopsin which acts like salivary amylase to hydrolyze starch first to dextrin and then to maltose Amylopsin has maximal activity at pH 7 0 A maltase capable of splitting maltose into glucose has been identified in pure pancreatic juice

Fats are subjected to the action of pancreatic lipase or steapsin which hydrolyzes neutral fat into glycerol and fatty acids its maximal activity is at pH 8 0 This digestion of fat is considerably enhanced by the concomitant action of bile salts on the duodenal contents Bile salts lower the surface tension between water and neutral fat so that the fat can become emulsified into minute globules Thus the surface area of fat available for enzyme activity is increased greatly and the rate of hydrolysis is accelerated The fat splitting power of pancreatic juice is trebled by the presence of bile salts Fat also aids its own digestion Liberated fatty acids unite with alkali in the bowel to form sodium salts or soaps that have the same emulsifying action as the bile salts

Regulation of secretion In the intact animal nervous and hormonal mechanisms combine to

gastroileal reflex The reflex stimulates propulsive motility in the ileum which then empties its contents into the colon in preparation for new material The ileocecal junction acts as a valve so that cecal contents cannot flow back into the ileum

Absorption Under normal conditions absorption of the products of digestion occurs exclusively in the small intestine The smaller molecules formed in the digestion of protein fat and carbohydrate enter the blood in one direction only whereas there is a constant flux of salts and water in both directions across the intestinal membrane The gastrointestinal membrane like all other membranes in animals is impermeable to water soluble solutions whose molecular weight exceeds 100 so that for these active processes rather than simple diffusion are required

Absorption of protein Proteins are absorbed as amino acids Immunologic studies suggest that whole proteins may be absorbed in tiny quantities but that any protein which passes the intestinal membrane as such does so in such small amounts that it is of no nutritional significance The intestinal mucosa has an active specific mechanism for the absorption of water soluble amino acids the full nature of which is not understood at present

Absorption of fat Glycerol and fatty acids are the end products of hydrolysis of most fats and the majority of digested fat is absorbed as fatty acid Almost all fat is absorbed into the intestinal lymphatics which eventually drain via the thoracic lymph duct into the venous blood Little fat is absorbed directly into the portal blood stream The reason for this bypass of the liver is not understood More than 90% of the fat in the intestinal lymph is triglyceride or neutral fat which indicates a rapid resynthesis of fat immediately after absorption Fatty acids pass into the lymph stream more rapidly than glycerol

It has been suggested that perhaps some neutral fat can be absorbed in a finely emulsified form Fat particles less than 0 5 microns in diameter can be found in the intestine but the degree to which these may be absorbed directly is not known

Absorption of carbohydrate Monosaccharides are carried across the intestinal mucosa The rate of absorption varies from sugar to sugar glucose and galactose are most rapidly absorbed The presence of phosphates in the mucosal cells led to the phosphorylation hypothesis to explain the transport but full understanding of the mode of transfer of monosaccharides across the mucosa has not been reached

Absorption of water and electrolytes The various digestive glands add large quantities of electrolytes and water to the intestinal contents Water and most electrolytes pass continuously in both directions across the mucosa of the small intestine The rates of movement both into and out of the duodenum and uppermost segments of jejunum are rapid and about equal This portion of the small bowel acts as an equilibrator that is it brings the contents delivered to it from the stomach into approximate pH and osmotic equilibrium with blood Neutralization dilution or concentration is required The versatile duodenum accomplishes this quickly Digestion then proceeds promptly and soon absorption is the predominant function The further the chyme progresses down the small bowel the more absorption rather than exchange or equilibration occurs A net gain to the body becomes the order of the region and then water and electrolytes leave the intestinal contents to enter the blood stream more rapidly than they pass into the contents The contents however remain nearly isoosmotic with blood even in the terminal ileum so that the removal of much of the water and most of the electrolytes by now mostly sodium chloride is left for the great dryer of the gastrointestinal tract namely the colon [C F C I D A J]

Colon The colon or large intestine may be described as a reservoir where water and food material that have escaped digestion and absorption in the small intestine may be stored and concentrated until the time for elimination In general this reservoir is composed of an expandable muscular sac attached to the lower end of the small intestine At the point of union is the cecum a blind pouch with an appendix attachment in some animals The rectum a short muscular tube continuous with the lower end of the colon affords an outlet to the outside The architecture and function of the colon vary widely among species In herbivores (grass eaters) the colon is relatively long and sacculated at its upper end and the cecum is large and long for the purpose of storage and digestion of cellulose compounds Carnivores (meat eaters) which have practically all digestion and absorption in the small intestine have a short muscular colon with a small and nonfunctional cecum The colon of the omnivores shows combinations of the structural units of the herbivores and carnivores Man does not have a particularly long cecum but he does have a sacculated upper colon allowing for the storage concentration and continued digestion of vegetable compounds which make up much of his normal diet (Fig 25)

Although there is no evidence of any secretion of digestive enzymes by the colon those enzymes that pass from the small intestine with the undigested food materials into the colon continue digestion for a time

The colon and cecum of all mammals provide an excellent environment for the growth and development of bacteria These bacteria act effectively on protein and carbohydrates that have escaped digestion and produce a number of volatile compounds and gases that may be either absorbed or excreted They can also synthesize certain vitamins especially the B vitamins

The principal secretion of the cells of the walls of the cecum colon and rectum is mucin which lubricates the accumulating feces and thus counteracts the loss of water that is absorbed from the colon The normal water content of the passed feces varies considerably in different The

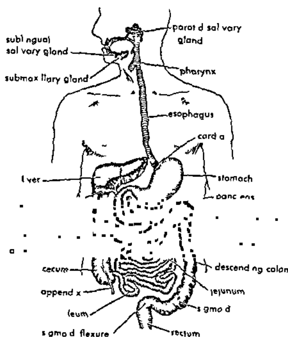


Fig 25 Alimentary tube and its appendages

average water content of the feces from the cow is approximately 83% whereas the fecal pellets of the rat may be less than 50% water Likewise the volume and weight of feces passed per day vary from animal to animal and are related to the amount of undigestible cellulose consumed It is estimated that the cow produces daily 60 lb of feces per 1000 lb of body weight while a 175 lb man on a mixed diet seldom produces more than 1 lb per day The composition of fecal material varies with the nature of the diet consumed and the completeness of the digestion and absorption in the small intestine Such substances as mucin live and dead bacteria and sloughed off epithelial cells make up a considerable percentage of the feces

For effective absorption of water and certain end products of digestion it is necessary that the contents be . . .
walls of
concentr
ward the rectum where it can be temporarily stored and finally expelled

Movements of the colon In general four types of movements take place in the colon to allow for the changes . . .

. . . ward or oral in direction These movements are not very powerful and have not consistently been observed in all mammals They have however been . . .
wh . . .
te . . .
colon contents and to aid absorption of water and digestive products into the blood stream

Segmenting or kneading movements These movements consist of a rhythmic contraction and relaxation of the colonic muscle without much change in position of the colon contents either forward or backward They are often very intense and their pumping action can be observed in several regions of the colon at the same time They knead the colonic contents and move the absorbable material to the lining of the colon where absorption into the blood stream occurs These movements are more predominant in the upper colon than in the lower when present in the upper colon they are absent from the lower, and conversely

Peristaltic movements These pushing movements are often observed after antiperistaltic and segmenting movements have been present They move material from the cecum and upper colon downward toward the lower colon and rectum for elimination Thus room is made available for more material from the small intestine for concentration and absorption

Mass movements This type of movement is seen only occasionally probably occurring not more than once every hour or two It is a prolonged powerful contraction of large areas of the colon at one time which slowly forces and packs fecal material into the lower colon often causing defecation

Nervous control of the colon The effect of the nervous system on colonic motility has not been exactly defined However, its anatomical presence cannot be denied

Enteric nerve plexus The enteric nerve plexus located underneath the mucosal layer and between the muscle layers is involved in the tonic and peristaltic activity of the colon In the absence or malfunctioning of the enteric nerve plexus marked distention of the colon and long periods without defecation may occur A case has been reported in which a man has lived more than one year without defecation and suffered no ill effects

Extrinsic nerve supply The extrinsic nervous supply to the colon is part of the autonomic nervous system and coordinates the colon with other parts of the gastrointestinal tract as well as with the brain and spinal cord These nerves can change the tonic and peristaltic activity of the colon reflexly and initiate movements of the colon involved in the act of defecation The gastrocolic reflex is an illustration in which the extrinsic nerve supply coordinates the stomach with the colon In this case the entrance of food into the stomach will within a few seconds set up increased motility in the colon which may last for several minutes and often causes the desire to defecate after the meal

The desire to defecate is first initiated by the forcing of fecal material into the rectum The nerves in the distended rectum stimulate sensory nerve endings and send sensory impulses into the spinal cord and on up to the hypothalamic area in the brain There connections are made with motor fibers that return via the spinal cord to the lower colon to produce a mass contraction and a relaxation of the anal sphincters and the expelling of feces It is well known that the act of defecation

can be temporarily inhibited by willful contraction of the external anal sphincter, and enhanced by abdominal contractions

Research development. Research and scientific knowledge of the colon in fact of the entire gastrointestinal tract, has lagged far behind advances in other fields of physiology Two obvious reasons may be cited lack of knowledge concerning the intimate nature of the contraction mechanism of smooth muscle fibers, and the extreme depressant effects of anesthetics on the normal tone and activity of the smooth muscle of the entire gastrointestinal tract.

The recent work of E. Bülbring and colleagues on the recording of electrical changes from single smooth muscle cells of the colon during stretched and unstretched states, as well as that of G Burnstock and R W Straub who made similar studies with different ions and drugs, indicates clearly that an interest has been created and basic contributions to knowledge at the cellular level are forthcoming

The techniques available for studying motility of the gastrointestinal tract point to the necessity for methods of study in both man and animals in the normal and unanesthetized state Although considerable information has been obtained by the use of x rays this technique does not afford much opportunity for studying actual pressure changes and absorption states in the normal functioning colon

Recent development of the Miller Abbott tube for sampling intestinal contents, and the telemetering capsule of J Farrar for recording pressure changes along the gastrointestinal tract shows great potentialities for obtaining accurate information on how the colon in man and other mammals normally functions [F R S]

Bibliography W Andrew *Textbook of Comparative Histology*, 1959, L B Arey, *Developmental Anatomy*, 6th ed, 1954, B P Babkin *Secretory Mechanism of the Digestive Glands*, 2d ed 1950, C H Best and N B Taylor, *The Physiological Basis of Medical Practice*, 6th ed 1955, E Bulbring and I N Hooton *J Physiol*, 125 292–301 1954, G Burnstock and R W Straub, *J Physiol*, 140 156–167, 1958, W M Copenhaver Endodermal derivatives, in B H Willier P A Weiss and V Hamburger (eds) *Analysis of Development*, 1955, H H Dukes, *The Physiology of Domestic Animals*, 7th ed, 1955, J T Farrar V K Zworykin and J Baum, Pressure sensitive telemetering capsule for study of gastrointestinal motility *Science*, 126 975–976 1957, M I Grossman, Gastrointestinal hormones *Physiol Rev*, 30(1) 33–90 1950, *J Clin Invest*, 36 1521, 1957, L L Langley and E Cheraskin, *The Physiological Foundation of Dental Practice*, 2d ed, 1956, A A Maximow and W Bloom, *A Textbook of Histology* 7th ed, 1957, O E Nelsen, *Comparative Embryology of the Vertebrates*, 1953, C L Prosser (ed), *Comparative Animal Physiology*, 1950, T A Rogers The metabolism of ruminants, *Sci American*, 198(2) 34–38, 1958, A S Romer *The Vertebrate Body* ed, 1955, D Rudnick, Development of the digestive tube and its derivatives *Ann N Y Acad Sci*, 55(2) 109–116, 1952, B T Scheer, *Comparative Physiology*, 1948, J B Sumner and K Myrback (eds), *The Enzymes*, vol 1 pt 1, 1950, C K Weichert, *Anatomy of the Chordates*, 2d ed, 1958, E M Vaughan Williams, *Pharm Rev*, 6 159–190, 1954, J Z Young, *The Life of Vertebrates*, 1950, W D Zoethout and W W Tuttle, *Textbook of Physiology*, 13th ed, 1958

Digital computer

Any device for performing mathematical calculations on numbers represented digitally, by extension, any device for manipulating symbols according to a detailed procedure or recipe *See* COMPUTER

The class of digital computers includes conventional adding and calculating machines and some data processing systems, as well as special purpose digital control devices for industrial processing and manufacturing operations, air borne devices, and the like *See* CALCULATING MACHINES, DATA PROCESSING SYSTEMS SAMPLED-DATA CONTROL SYSTEM

In this article emphasis is placed on large scale stored-program digital computers These machines store internally several thousands of numbers or other items of information and control and execute complicated sequences of numerical calculations and other manipulations on the stored information without human intervention and according to instructions also stored in the machine

Computation consists of a sequence of operations on operands or data The specification of an operation is an instruction, and a list of instructions to be followed in operating on data is a program Thus, for example, the computation called for in evaluating $d/a(b + c)$ with a desk calculator and pencil and paper would be a sequence of operations on a, b, c, and d (the data) specified by a list of instructions (the program) including directions such as "multiply $(b + c)$ by a and copy the result appearing in the register onto paper in region 1"

All programs include besides arithmetical commands such as add, instructions for copying numbers into or out of machine registers and for writing or storing ("remembering") data in some medium Programs also include instructions for causing computation to jump, or branch that is, to depart from one course of computation when a result in a register has some specified property (such as being negative) and to enter another course

A stored program digital computer is a device with input and output means for accepting instructions and data and printing or otherwise presenting results to the human user, an internal memory, or storage device for storing instructions awaiting execution, recording intermediate results and accumulating final results for output, an arithmetic unit with various registers for holding oper[illegible] devices for adding, subtracting, multi

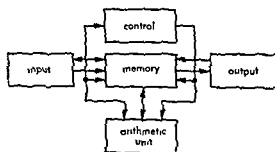


Fig 1 Schematic diagram of computer organization

dividing shifting and other operations, and a control unit which selects interprets and actuates the execution of instructions according to the programmed specifications which are stored in the memory (see Fig 1)

All data and coded instructions (represented by groups of binary or decimal numbers) are initially stored in the memory of the computer at specific locations each location having a numerical address associated with it During computation both instructions and data are referred to by their addresses which either remain fixed during a computation or are changed by a program which transfers information to new locations

The rest of this article discusses the logical structure and hardware employed in digital computers For a discussion of programming see DIGITAL COMPUTER PROGRAMMING. For a discussion of input and output devices see DATA PROCESSING SYSTEMS

Digital representation Digital computers operate with numbers represented internally by two-state devices described later in this article Accordingly radix 2, or binary number systems and arithmetic are employed in digital computers For detailed discussion of binary numbers see NUMBER SYSTEMS For reference Table 1 lists the first twenty binary numbers and their decimal equivalents

Table 1 Binary-decimal equivalents

Binary	Decimal	Binary	Decimal	Binary	Decimal	Binary	Decimal
0	0	101	5	1010	10	1111	15
1	1	110	6	1011	11	10000	16
10	2	111	7	1100	12	10001	17
11	3	1000	8	1101	13	10010	18
100	4	1001	9	1110	14	10011	19

From the point of view of the user particularly in data processing computers it is desirable to manipulate decimal and alphabetical data directly In some computers therefore data are represented by coded groups of binary digits Table 2 shows three typical binary codes the numeric 8-4-2-1 and 2-out-of-5 codes, and the alphanumeric excess 3 code

Using such schemes, decimal or alphabetic characters are written as sequences of code groups Thus 268 in the 8-4-2-1 code is 0010 0110 1000, CAB in excess 3 code is 0010110 1010100 0010101

Table 2 Binary-coded decimal and alphabetical characters

Character represented	Numeric 8421	Numeric 2 out of 5	Excess 3	Character represented	Excess 3
0	0000	00011	1000011	I	0011100
1	0001	00101	0000100	J	1100100
2	0010	00110	1000101	K	0100101
3	0011	01001	1000110	L	0100110
4	0100	01010	0000111	M	1100111
5	0101	01100	0001000	N	1101000
6	0110	10001	1001001	O	0101001
7	0111	10010	1001010	P	0101010
8	1000	10100	0001011	Q	1101011
9	1001	11000	1001100	R	0101100
A			1010100	S	1110101
B			0010101	T	1110110
C			0010110	U	0110111
D			1010111	V	0111000
E			1011000	W	1111001
F			0011001	X	1111010
G			0011010	Y	0111011
H			1011011	Z	1111100

8-4-2-1 code This has the added feature of being a weighted code because a decimal equivalent can be recovered by adding the weights corresponding to 1s in the code group Thus 0101 is $0 \times 8 + 1 \times 4 + 0 \times 2 + 1 \times 1$ or 5 Neither of the other two codes is weighted

2-out-of-5 code This code is self checking Each group contains exactly two 1s. Any odd (and some even) number of errors in transmission of a number so coded would yield a group with either more or less than two 1s. Such errors may be easily detected by circuits as described later in the article

Excess-3 code The alphanumeric version of the excess 3 code so-called because the binary value of the four low order digits is three in excess of the decimal it represents, is an example of a self-complementing code The nine's complement of the numeric part of the code group (four low-order digits) is obtainable by reversing the digits Thus the nine's complement of 0100 (decimal 1) is 1011 (decimal 8) as required This code is convenient for complementation but somewhat complicated for other arithmetical operations

The excess-3 code as depicted also includes in the high order position a redundant parity bit which is selected in such a way as to make the sum of the 1s in the code group odd This makes any single error in code transmission easy to detect as the error yields a group with an even number of 1s

Use of codes An ordered collection of symbols or characters is a *word* Examples are 11011 CAB and 0011 1000 1001 In the first example each character is a bit or binary digit (see BIT), in the second an alphabetical symbol and in the third a (binary coded) decimal digit

Storage locations and the various arithmetical registers in a computer usually accommodate words of a standard fixed length Most binary machines accommodate words of 35 or more bits

most decimal machines accommodate 11 or 12 decimal or alphabetic characters (including algebraic sign) Small numbers are represented by words with the high order positions filled with zeros Thus in a 10-decimal digit word 468 appears as 0000000468 and in a 16-bit binary word 11011 appears as 0000000000011011 Digital computers operate with words corresponding to data or instructions as described earlier A word representing a datum is a datum word a word representing an instruction is an instruction word

Assume a simplified computer SC in which words consist of six decimal digits including sign Data words contain a left most sign position (+ or −) and five numerical digits for example +00278 An instruction word consists of a sign (always +) two operation code digits and the address of an operand For example the word +01279 means add to the accumulator the number stored in memory location 279 01 in particular means add in SC

If such words were to be coded in the 8 4-2 1 system +00278 would be represented by

+0000 0000 0010 0111 1000

and the instruction Add to the accumulator the number stored in memory location 279" by

+0000 0001 0010 0111 1001

In transmitting data within a computer words may be sent a bit at a time over a single channel low order bits or digits first in which case transmission is serial by bit In the examples transmission would therefore require 24 bit times (including sign which has not here been coded)

Alternatively words may be sent over four channels (or more depending on the code) one for each bit in a code group

(	0 0 0 0 1	(8 bit channel)
+	0 0 0 1 0	(4-bit channel)
	0 0 1 1 0	(2 bit channel)
)	0 0 0 1 0	(1 bit channel)
	0 0 2 7 8	*(decimal equivalent)*

It is customary to think of transmission as occurring from left to right in time (alternatively stated time moves left) According to this scheme transmission is said to be parallel by (decimal) digit or by (decimal or alphabetic) character Transmission in the example would require only six bit times On binary machines words are transmitted either serially or parallel by bit Binary or decimal machines are termed either serial or parallel depending on the *nature of the information transmission* On magnetic tape binary words may be partitioned arbitrarily into groups (such as six 6 bit groups in a 36 bit word) for multichannel transmission

states or by a sequence of high or low voltages occurring in transmission channels or in switching

Fig 2 Waveform representation of binary number 101101 by series of (a) voltage levels (b) voltage pulses

networks In Fig 2*a* the binary number 101101 is represented serially by a sequence of voltage levels and in Fig 2*b* by a sequence of voltage pulses In Fig 2*a* the dc signal is held positive throughout the time allotted for a bit if 1 is being represented and relatively negative or zero if 0 is being represented In Fig 2*b* 1 is represented by a pulse and 0 by the absence of a pulse In a similar way the parallel by digit train could be represented by four parallel sequences of waveforms of either kind More details concerning generation and transmission of signals and problems of timing appear later in the article

Switching and memory devices These are used to perform arithmetic and control operations in digital computers All such devices embody certain primitive logical and memory functions some of which are treated as black boxes in Fig 3 Figure 3*a* represents a logical AND function Each of the inputs and the output are capable of carrying a 1 or a 0 signal The output of the AND device is 1 only when both the input *A* AND the input *B* are 1 Figure 3*b* shows the logical OR function The output is *1 when either A* OR *B (or both) inputs are 1* The single input box in Fig 3*c* is a logical NOT function The output is 0 if *A* is 1 and 1 if *A* is 0 The names of these functions are taken from the theory of the propositional calculus in mathematical logic Applications of logic to switching and computer circuits are found in both circuit analysis and design *See* LOGIC SWITCHING THEORY

Fig 3 Representation of logical functions (a) (b) OR (c) NOT (d) Memory

Figure 3*d* shows a memory function In a sense it remembers its input signal As an example suppose that signals 0 or 1 occur on inputs A and B at bit times 0 1 2 3 n If A is 1 at time t then the output is 1 at $t+1$ and subsequent times if B is 1 at t then the output is 0 at $t+1$ and thereafter if neither A nor B is 1 at t the output at $t+1$ remains in whatever state it was in at time t Such devices are also called state devices and are bistable (state 0 or state 1) in character

Many computer registers are constructed of sets of devices having the function of Fig 3*d* Clearly two such devices can represent four binary numbers (00 01 10 11) where the output of each of the two elements is a bit In general n binary numbers can be represented by a multiplicity of state devices equal to the least integer equal to or greater than log n which is of course the same as the number of bit positions in the number

Various realizations of these and other logical functions are possible and the types used depend on the operating speeds reliability and cost considerations specified for the computer In the following relays vacuum tubes and germanium rectifiers are discussed as typical realizations of the functions described above Transistors as regards logical performance are handled in a way quite analogous to vacuum tubes but the actual circuit requirements are somewhat more complicated For complete discussions of transistor switching and memory devices together with discussion of the use of magnetic cores cryotrons and other state de

rences

typical switching relay When lead A is grounded after a

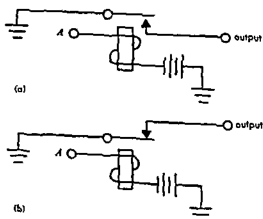


Fig 4 Typical switching relays (a) Normally open contacts (b) Normally closed contacts

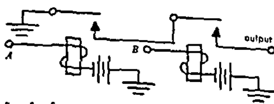


Fig 5 Series circuit providing logical AND function

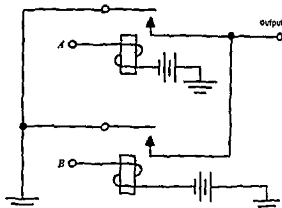


Fig 6 Parallel circuit provides logical OR function

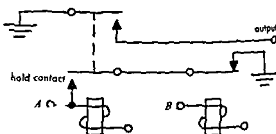


Fig 7 Relay memory device

small operate time (ranging from a few to perhaps 100 msec or more) the contact closes and the output is grounded (= 1 logically) Such a contact is called normally open

In Fig 5 the contacts on relays A and B are connected in series thus providing the AND function there is an output signal when both A AND B are closed (= 1)

The parallel hookup of the contacts in Fig 6 provides the logical OR function because an output occurs when A OR B or both are closed

In Fig 4*b* there is an output only when A is not grounded therefore the device realizes the NOT function The output is 1 when A is not grounded and 0 when A is grounded The working contact i said to be normally closed

Figure 7 shows a one-bit two state memory using a relay with a hold contact When input A is grounded both contacts on the relay close and are held closed by the ground supplied through the hold path even when A is ungrounded To release the relay a normally closed contact in the feedback path opens when relay B is energized

This device remembers at $t+1$ and subsequently a signal on A at t or a signal on B at t Consequently the output is precisely the logical memory function of the inputs abstractly defined previously It is assumed that A and B never are grounded at the same time t

Figure 8 shows how switching networks on relay contacts can be built up to realize fairly complicated logical functions The circuit shown is an error detector for the 2 out of 5 code appearing in Table 2 Signals appear parallel by bit on the inputs A E Whenever the coded group of bits con

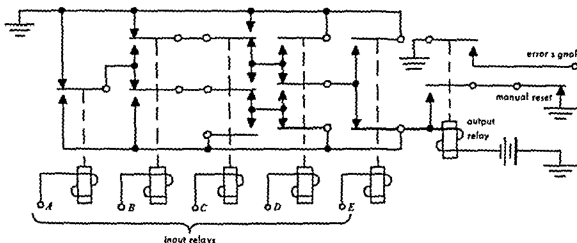


Fig 8 Relay error-detector circuit for 2-out-of 5 code Input relays are shown in unenergized position (for a 0 input) Output relay is energized by ground connection only when there are more or less than two 1 inputs Input relays are energized by a 1 input

tains more or less than two 1 inputs the circuit puts out an error signal which stays on until the device is manually reset There are 22 possible erroneous combinations such as 00000 00010 01101 11110 and 11111 which contain other than two 1 inputs Each input lead receives either a 1 or a 0 input A for the first digit position B the second and so on, thus for an input 10000 A receives a 1 input (ground) and B C D and E all receive a 0 input (no ground)

Hereafter NOT A will be abbreviated by $\bar{A}$ A OR B by $A + B$ and A AND B by AB

Relays are quite reliable in action and fairly inexpensive but they are relatively slow Although computers have been made with exclusively relay logic relays are now restricted to controlling the mechanical card punching tape driving and printing operations associated with computer input and output

Vacuum tube switching Electronic switching circuits usually use vacuum triodes although multigrid tubes and gas tubes are also in common use The inverter circuit of Fig 9 consists of a triode with input to the grid through a voltage divider R_1 and R_2 are chosen in such a way that an input signal either holds the grid at ground (or slightly above) so the tube conducts or low enough negatively to cut the tube off entirely The two levels of the input A may be taken as V_1 and V_2 so that V_1 is the same as the plate supply and V_2 the voltage at the plate output when the tube is conducting Calling $V_1 = 1$ and $V_2 = 0$ if the input is 1 the output is 0 and vice versa which realizes the NOT function

The OR function may be obtained by the twin triode cathode follower of Fig. 10 If positive signals (logically 1) are applied to either A or B or both the cathode follower output is 1

The AND function is not easy to represent directly however a possible arrangement using 4 tubes is given in Fig 11 There is a high c

the common load resistor R_L only when both tubes are in a cut-off condition that is the output is 1 only when points X and Y are logically 0 However X and Y are logical NOT functions of A and B respectively Hence the output is 1 only when A and B are both 1 which is precisely the AND function desired

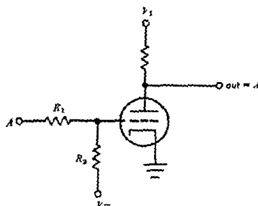


Fig 9 Vacuum-triode NOT circuit

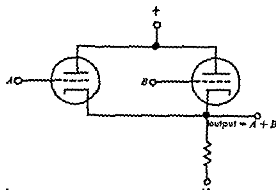


Fig 10 Vacuum-tube OR circuit

Fig 11 Vacuum tube AND circuit

The memory function is achieved by a flip flop which is an assembly of two *inverters* as in Fig 12 If tube T_1 is cut off *the potential at the grid of* T_2 is high enough so that T_2 conducts In this situation the output *at X is relatively high* or 1 and that at Y is low or 0 Now if *a pulse* is applied through the capacitor *at A the device* flips abruptly T_1 then conducts and T_2 *is cut off* and the logical values of *X and Y* are reversed *If B* is *pulsed the* device resets to the first state described Thus the device is bistable and *will remain* in ei ther state ... the X output 1 outp... ... schematic memory device and $X = 0$ $Y = 1$ *to a 0 output* Consequently in summary we may say that as in the earlier figure the output *is* 1 *at* $t + 1$ when *A* is 1 at *t* and is 0 at $t + 1$ when *B is* 1 *at t X will* hereafter be called the ON output and *Y* the OFF (see Fig 15)

Tube AND OR and NOT *devices* operate practically instantaneously Flip flops *can be designed* to change state within a matter of millimicroseconds Consequently the tube *logic is conducive* to greater computer speeds than the relay type *Basic* bit *time* rates of more than a megacycle *per second* ...

... space consumption lower cost and easy adaptability to printed circuit and other packaging techniques

Figure 13*a* shows a diode rectifier AND circuit Each diode conducts in the direction of the arrow and has essentially zero conductivity in the other direction If either *A* or *B* is at low potential relative to $V+$ current flows through the diode and the voltage drop across the resistor keeps the output at a low potential (0) However when both *A* and *B*

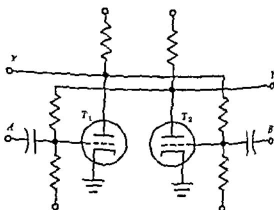


Fig 12 Vacuum tube memory circuit (flip flop)

are high (logically 1) there is no current and consequently the output is high (1) This is by definition the AND function

If an analogous manner the OR function is obtained by the device of Fig 13*b* If either *A* or *B* is at high potential the output is high

An alternative AND device for gating pulse signals ... readily ... the out... AND and OR switches are sometimes called gates current usage tends to restrict the term either to circuits as

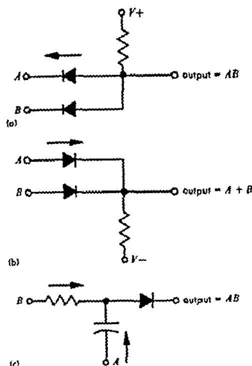


Fig 13 Diode switching circuits (a) AND circuit (b) OR circuit (c) Gated AND circuit

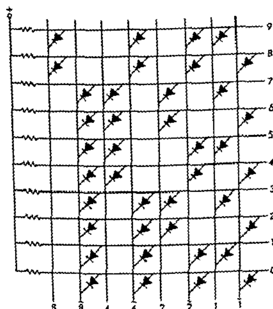


Fig 14 Diode matrix for decoding a binary 8 4 2 1 group into decimal form

in Fig 13c or to other AND circuits in which one of the inputs is of a pulse tube

Figure 14 shows a diode matrix that can convert an 8 4 2 1 binary coded group into decimal form This employs ten AND gates each having four inputs The output leads at the right receive a 1 output when all four diodes are cut off Each input has two possible conditions for example 8 or $\bar{8}$ Thus to obtain an output at 9 input pulse must be received on the 8 $\bar{4}$ $\bar{2}$ and 1 input leads

Switching and memory assumptions For brevity's sake certain assumptions will be made for subsequent discussion of arithmetic and control circuits

1 The basic constituents are diode AND and OR circuits tube inverters, and flip flops

2 Significant information occurs only at times $t = 1\ 2\ 3$ defined by pulses generated by a clock (electronically a multivibrator) these may be taken as square pulses of ½ μsec duration ½ μsec apart

3 All detailed hardware problems such as amplification wave shaping, and impedance matching have been taken into account permitting arbitrary logical arrangements of the basic circuit elements

4 Flip-flops require pulse type inputs

The various logical and memory devices will be schematically represented as in Fig 15 Although assumption (1) is used in this discussion relays transistors magnetic cores and cryotrons may all be used more or less indifferently as far as computer logic is concerned This statement is not necessarily true of possible future computers which may use logically complex devices decomposable by control means into various functions

Concerning assumptions (2) and (4) the example of Fig 16 illustrates the type of timing technique assumed Switching occurs between clock pulses Thus before $t = 1$ the signal on input A is assumed to have been applied It reaches its high dc level at 1 Because B is 0 AB equals 0 and S equals 0 at this time

At $t = 2$ both A and B are 1 so $AB = 1$ and through the second gate coinciding with the clock pulse $S = 1$ By $t = 3$ M has been flipped by the pulse on its set input and its ON output is 1 At $t = 5$ $R = 1$ and by $t = 6$ M is reset to 0 (OFF is 1) The not too unrealistic assumption is

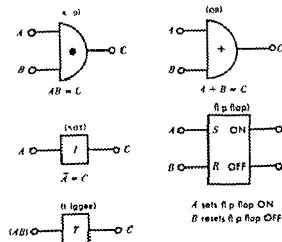
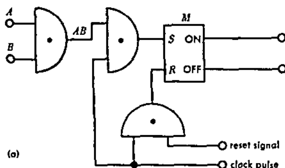


Fig 15 Schematic representation of logical and memory devices A and B are inputs C is the output the A and B inputs to a flip-flop are tied is call As pulses (A B) enter C



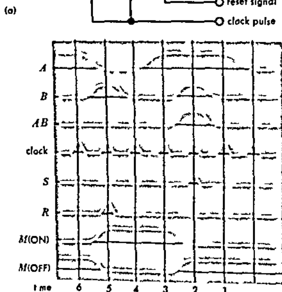


Fig 16 Timing technique (a) Waveforms (b) Circuit

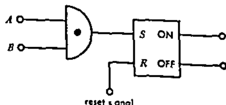


Fig 17 Simplified circuit of Fig 16

made that AND OR NOT switching take equal times

Having acknowledged assumptions (2) and (4) the circuit of Fig 16 may be simply drawn as in Fig 17

A computer in which information is defined by a train of clock pulses as above is synchronous One operating in such a way that a given sequence of operations may commence as soon as a prior one terminates without superimposed clock timing is asynchronous All the following discussion is adaptable to the asynchronous case by proper design of the switching devices The asynchronous type is inherently capable of faster operation than the synchronous

Arithmetic devices The devices for performing addition subtraction and multiplication are discussed in this section for a hypothetical computer having 10-bit binary words which are transmitted in parallel Typical schemes will be discussed except that for simplicity it will be assumed all numbers are positive that only those subtractions and divisions occur for which there are positive results and that no answers overflow the registers These restrictions are of course unreasonable for actual design but the basic principles can be best discussed under them Appreciation of the variety of systems possible can be gained from the references listed in the bibliography

The arithmetic unit contains three 10-bit flip-flop registers as in Fig 18 Only the circuitry associated with bit positions A_1 and B_1 is shown in any detail It is assumed that two 10 bit words are represented at the outputs of the memory register and the accumulator (for the time being ignore the X register) It is required to add the numbers together and put the sum back into the accumulator The subscripts on A and B at the outputs of the register flip flops and the inputs to the adder indicate the order of the bit

Addition and subtraction Binary addition for two single bit numbers is defined by Table 3 A and B stand for the corresponding bits in the addend and augend at the ith order, S_i stands for the sum and C the carry The table shows that when A and B are 0 both the sum and the carry are 0 when either A or B is 1 and the other is 0 the sum is 1 and the carry is 0 $(1 + 0 = 1)$ when A and B are both 1 the sum is 0 and the carry is 1 $(1 + 1 = 10)$

Table 3 Binary addition table

A	B	S	C_i
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

A half adder performing single bit binary addition of inputs on A and B and verify the S and C outputs The sum output occurs only when A or B but not both is 1 This logical function is called an exclusive OR and the circuit (excluding the carry of Fig 19) a quarter adder

A full parallel adder may be obtained by connecting half adders together In Fig 20 this is done for three stages or orders to exemplify the procedure For ten stages this procedure specifies the design of the entire adder of the arithmetic unit of Fig 18

For illustration in Fig 20 the numbers (A) 110 and (B) 011 are added the low order digits correspond to k and the high to $k + 2$ in the diagram As indicated in parentheses at the outputs the sum is 1001

Subtraction may be accomplished either directly or by complementation and addition In the latter method subtraction is done by adding the ones complement of the subtrahend to the minuend and then adding 1 to the lower order position of the sum (the ' end around carry) The one s complement is

Fig 18 Arithmetic unit

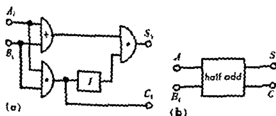


Fig 19 Half-adder (a) Circuit schematic (b) Block diagram

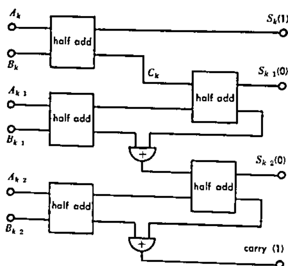
obtained by simply interchanging 0s for 1s and vice versa in each bit position of the entire subtrahend word, or, by circuit means, by sending the bit through a quarter-adder, the other input of which is held 1 during a bit time To illustrate the method 011101 is subtracted from 100000 The assumption is made that the registers are of six bit capacity

Direct subtraction		*Complementing method*	
100000	(decimal 32)	100000	
011101	(decimal 29)	100010	(one's complement)
000011	(decimal 3)	000010	(sum)
		1	(end-around carry)
		000011	(difference)

To use the previously designed adder for subtraction, all that is required is to provide means for complementing each bit in the memory register (which holds the subtrahend) and for adding unity to the low order position of the adder when using it for subtraction This entails a special design of the first stage of the full adder This is given in Fig 21

When $K = 0$, this circuit acts like the half adder with outputs S_k and C_k of Fig 20 However, when $K = 1$, the circuit acts like a single bit position adder for three bits as is required for the complementing method

To see how addition takes place, assume that the augend B_1 is in the accumulator and the addend A_1 in the memory register of Fig 18 An add pulse at $t = 1$ from the computer control (described later) sets the flip flop marked + ON by time $t = 2$, this output holds open the AND gates (illustrated at positions A_1, B_1 in Fig 18) at the outputs of all stages of the two registers At the same time the output from the + flip flop resets all flip-flops in the accumulator so that by $t = 3$ the accumulator is cleared and ready to receive the sum S_1 S_{10}. The sum, delayed one bit time (delays are not shown in the diagram) appears on the set inputs to the accumulator at $t = 3$ and by $t = 4$ appears in the accumulator Also, at $t = 2$ a reset p from the control clears the + flip-flop to



	$k+2$	$k+1$	k	
(A)	1	1	0	
(B)	0	1	1	
(A + B) 1	0	0	1	(sum)

Fig 20 Three stages of full adder

For subtraction the minuend is in the accumulator and the subtrahend in the memory register An add and a complement pulse from the computer control set + and K flip flops respectively The operation then continues as in addition except for the following (1) the quarter adder at the input to the A_1 A_{10} ouputs complements the A bits because the K flip flop is ON (2) the K input to the first stage of the adder is now 1 and effectively adds the end around carry needed in subtraction (Fig 21) (3) the K flip flop is also reset to OFF at $t = 2$ The timing pulses $t = 1$ 2 used above require a counter which is not shown *See* COUNTER DIGITAL

...nes the rate at which a computer can go through computations On the schematic computer of Fig 18 the logic used for multiplication is perhaps the most obvious type but in actual hardware implementation is very satisfactory as regards speed The scheme is one which is executed by over and-over addition In the example of Fig 22 101 11 is the multiplicand and 110 01 the multiplier The multiplicand is placed in the memory register and the multiplier in the X register (often called the multiplier quotient register) The accumulator is reset to zero The procedure is to examine the low est order bit in the X register If it is 1 the multiplicand is added into the accumulator and then the contents of both the accumulator and the X register are shifted one place to the right If the lowest-order bit appearing in the X register is 0 the contents are simply shifted This process is repeated 10 times (or in general a number of times equal to the bits in the standard word) and the product then appears in the X register and extends into the accumulator It is evident that a precision of 20 bits in the answer is possible because the two registers act as one double word length register As seen in the example the decimal point must be computed by the human user and is placed four places to the left of the right hand end of the X register The references should be consulted for further discussion of point fixing scaling methods and floating point arithmetic

Assuming control and timing means which are not treated here the arithmetic unit is entirely adequate for multiplication provided the accumulator and X register are designed so as to shift right on receiving a shift pulse A section of the accumulator is shown in Fig 23 with means supplied for shifting the ith bit into the $i-1$ flip-flop During multiplication a source of 10 pulses would have to be provided for the shift phase of the add shift cycle A register thus equipped is a shifting register

Second during the add phase if flip flop X_1 of the X register is 1 the + flip flop must be set This objective is gained by leading the ON side of X_1 into the OR gate at the + flip flop of Fig 18 A pulse at the lead indicated which must come between each of the 10 shift pulses is then provided Both the shift and add pulses must be far enough apart in time to permit the + flip flop to flip the accumulator to be cleared and finally to receive the sum of the multiplicand and previously formed partial product

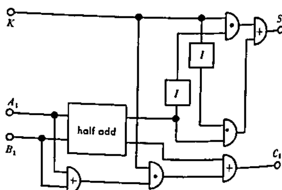
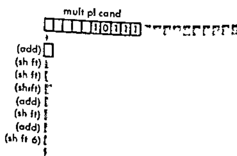
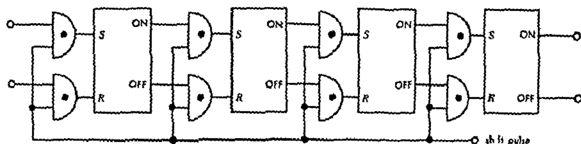


Fig 21 First adder stage providing for end-around carry



(101 11) × (110 01) = 100011 1111
(5 75) × (6 25) = (35 9375)

Fg 22 Mult pl cat on by over-and over add t on

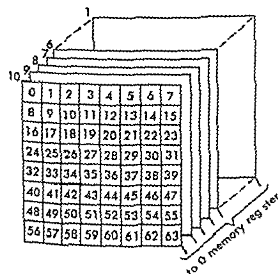


Fg 23 Use of sh ft pulse

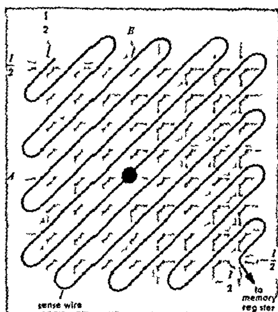
D vision Division is performed with this type of unit in a way analogous to multiplication but by way of 10 cycles of subtraction and left shifts The divisor is placed in the memory register the dividend in the accumulator and the quotient is then formed in the X register The scheme employed is much like that of successive subtraction with pencil and paper Details of the additional circuit requirements are omitted

Other operations Other arithmetical operations such as root extraction must be programmed as sequences of the basic arithmetic operations on almost all digital computers For more complicated mathematical operations such as integration and differentiation suitable numerical procedures must first be found and then programmed in terms of the given computer operation code For frequently employed mathematical operations subroutines may be written and incorporated in larger programs as needed

Memory devices A high speed *memory* with a capacity of 1000–32 000 (or even more) words used for storing data intermediate results of computation and program instructions may be designed using any one of a wide variety of available techniques For various memory or storage devices and systems *see* STORAGE DEVICES



Fg 24 Ten-plane 64 b ts per plane core memory



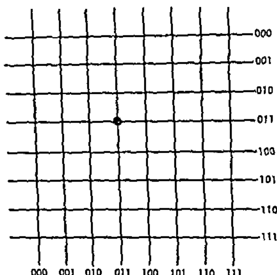
Fg 25 Matrix of cores

In this article the schematic arithmetic unit of Fig 18 is combined with a simplified 64-word immediate access magnetic core memory which implies that in the machine any 10 bit word may be directly selected by gating circuits parallel by bit

The core memory consists of 10 planes (one per word bit) of 64 bits each each stored word spans all 10 planes (Fig 24) The storage locations carry addresses 0–63 Each word is nine bits plus the sign so in memory the + or − (0 or 1) occupies the 10-plane and the lowest order bit the 1 plane

Figure 25 shows a core matrix which makes up a single plane Each core of the array has two windings (horizontal and vertical lines) When current is passed through both windings (in direction of arrows down and to the right) the magnetomotive force is sufficient so that the core goes to and stays in a state of flux 0 Current passed in the opposite direction (up and left) will change the state to 1 The core is thus bistable Only one core (and the corresponding core in every plane) is pulsed at a time The darkened core in Fig 25 for example is set in the 0 state by coincident currents on *A* and *B* but no other cores are so

The wire threaded through all cores is the wire (one f r each plane) When *A* ar



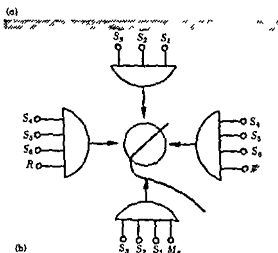


Fig 26 Address selection for core memory (a) Core layout (b) Gating

driven for example if the darkened core holds a 1 it will switch to 0 and send a pulse over the sense wire to the memory register (see Fig 18) setting a flip flop

To work properly computers must be so designed that data read from memory are still retained that is when read information is copied not erased Because reading is achieved as implied in the previous paragraph by switching the core to 0 *if the* core previously held a 1 it will be lost This is *known as destructive read out* In core memories of this type it is therefore essential that the addressed bit be regenerated Regeneration is accomplished in one possible way by writing *the contents of the* memory register just loaded by *the read operation* back into the addressed location Assume therefore that reading always takes place in two stages read and regenerate

Writing is done simply by addressing the appropriate horizontal and vertical *inputs on the* right and bottom of the matrix and *by driving current* into the selected core putting it in state 1 regardless of the previous state Information is written from the previously loaded memory register

Memory selection and reading and writing control are indicated in Fig 26 The manner of address selection is indicated for (as an arbitrary example) plane 6 In binary notation the 64 addresses range from 000000 to 111111 The left hand three bits select one of the eight horizontal wires and the right hand three bits select a vertical wire This should be compared with the address designation of Fig 24 and the core scheme of Fig 25 To illustrate the darkened core is located at address 011011 (decimal 27)

In Fig 26*b* the gating for selection of a bit is shown the letters S denoting outputs from an address selection circuit (Fig 27) R is a read signal W a write signal and M_6 designates an input from flip flop 6 of the memory register, this is the bit to be gated into the core

It is entirely possible to take information from cores into the memory register (including time for regeneration) using a core memory of this type in a few (10 25) microseconds Experimental machines (1959) are capable of access time shorter than 1 μsec

Digital computer control This includes all those computer elements required to interpret programmed instructions and to sequence and actuate machine computations in the order required A sketch of these matters may be approached by first considering the instruction list of the schematic computer given in Table 4 For explanations of these instructions and for additional instructions *see* DIGITAL COMPUTER PROGRAMMING

An instruction word always contains a + in the 10th bit position the Op code is located in positions 9 8 7 and the address part in 6–1

Table 4 List of computer instructions

Instruction	Op code	Address part
Clear add	+000	XXXXXX
Add	+001	XXXXXX
Sub	+010	XXXXXX
Load X	+011	XXXXXX
Mult	+100	XXXXXX
Store	+101	XXXXXX
Un Branch	+110	XXXXXX
Branch >0	+111	XXXXXX

The address register (Fig 27) is a six bit counter which holds the address of an instruction to be executed assuming that a coded program using Table 4 has been placed in the core memory For illustrative purposes suppose the following simple program is so stored sum a and b and if the result is > 0 put it in the location in memory whose *address is 100000, if however* the sum is $\leqq 0$ put it in 100001, to begin with a is stored in 010000 and b in 010001 the program is to be stored in 000000–000111 This program is coded in Table 5

The address register holds 000000 (instruction location of first instruction), and then sequences to

Fig 27 Example of computer control functions

Table 5 Coded program

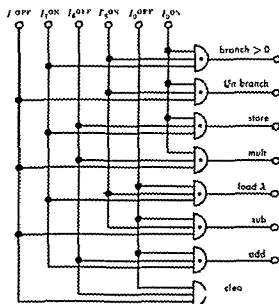
Instruction number	Instruction location	Instruction	Op code	Address part
1	000000	Clear Add	000	010000
2	000001	Add	001	010001
3	000010	Branch >0	111	000101
4	000011	Store	101	100001
5*	000100	Un Branch	110	000100
6	000101	Store	101	100000
7*	000110	Un Branch	110	000110

* Effectively stops computation

000001 000010 000110 moving the computer through the program (as will be described) unless a branch instruction is encountered

Whatever address is in the register is selected from the memory through the selection circuit (Fig 27) when an R pulse (Fig 26b) occurs This causes the contents (in the example the first instruction which is +000010000) to go through the memory register to the instruction register (In practical machines it would go directly not through the memory register) This complete operation is called an instruction selection cycle

Next the Op code of the instruction is decoded and the address part through positions 6–1 of the instruction register selects via the address selection circuit (through on circuits) the datum a (at 010000 in example) In Fig 28 the decoder is depicted The inputs I_9, I_8, I_7 are the outputs of the flip-flops of the instruction register positions 9 8 and 7 The word selected is read into the memory register and the decoded clear add signal (see Fig 28) causes the contents of the memory register to go through the adder and into the reset accumulator Meanwhile the address register has counted to 000001 and the instruction register is reset This ends an execution cycle Next instruction 2 located



28 Decoder

at 000001 is selected by the address register and so on

On the Branch > 0 instruction no 3 the instruction is sent to the instruction register as usual the Branch > 0 signal from the decoder causes the accumulator to be scanned for > 0 contents (details of procedure omitted) and if > 0 the address part of the branch instruction is sent to the address register for the next instruction selection cycle If the accumulator is not > 0 the instruction register is reset and the next instruction is taken from the address which normally sequences by 1 in the address register

The store instruction selects an address through the selection circuit for writing the contents of the accumulator in the memory location specified To this end the accumulator contents during the execution cycle are first transferred to the memory register and thence to memory

The foregoing discussion although limited to an example summarizes the operation of a digital computer in one quite typical form The questions of input and output and special purpose design

matics and Logic for Digital Devices 1958 C C Gotlieb and J N P Hume *High Speed Data Processing* 1958 E M Grabbe (ed) *Automation in Business and Industry* 1957 W S Humphrey Jr *Switching Circuits* 1958 M Phister *Logical Design of Digital Computers* 1958 R K Richards *Arithmetic Operations in Digital Computers* 1955 R K Richards *Digital Computer Components and Circuits* 1957

Digital computer programming

The art of writing instructions for the operation of a stored program digital computer or more generally for specifying the steps in the solution of a computational problem whether by hand or by an automatic digital or analog computer

In this article digital computer programming is discussed with respect to a simplified computer SC the exact mechanical and electronic nature of which is not specified Basic programming concepts are introduced relative to this model but these concepts apply in conformity with current usage to all stored program machines

SC contains a high speed memory (or *equivalently store*) for storing 1000 six decimal digit numbers

computer conceptions implicit in this formulation of programming see DATA PROCESSING SYSTEMS DIGITAL COMPUTER

A program consists of a sequence of instructions from Table 1 Because the computer operates on data previously stored in its memory and according to instructions so stored the programmer must provide in advance for storage locations or addresses (000–999 for the 1000 words) of both the program and the data For example, suppose one wishes to add +478 to +739 Then one may arbitrarily store +00478 in the memory location (whose address is) 010 and +00739 in 011 The answer is to be stored in 012

The program would then be

000 Clear the accumulator and add the number in location 010 into it,

001 Add the number in location 011 into the accumulator,

002 Store the contents of the accumulator in 012

Each of the three instructions is itself provided with an address (000 001, 002) Note that the addresses but not the data are referred to in the instructions

Table 1 SC instruction list

Instruction	Op code	Address part	Remarks
Clear add	00	XXX	Contents of Acc are replaced by number stored in memory location XXX number is copied (also *remains in* XXX)
Add	01	XXX	Number in XXX is added to contents of Acc, while preserved in XXX
Sub(tract)	02	XXX	Number in XXX is subtracted from contents of Acc while preserved in XXX
Load X	03	XXX	Contents of the X register (*multiplier-quotient register*) are replaced by contents of XXX which are also preserved in XXX
Mult(iply)	04	XXX	Contents of XXX are *multiplied* by number in X register and rounded left hand five digits of product appear in Acc
Div(ide)	05	XXX	*Number* in accumulator is divided by number at XXX and quotient appears in X register
Store	06	XXX	Contents of XXX are replaced by number in Acc number is preserved in Acc
Store X	07	XXX	Similar to store
Un(conditional) branch	08	XXX	*Normal* sequencing is broken and computer goes to XXX for its next instruction
Branch >0	09	XXX	Normal sequencing is *broken* if contents of Acc are >0 computer goes to XXX for next instruction *if zero or* minus computer follows normal sequence

SC is so designed that when the program is placed in memory and the computation started SC normally sequences itself from the first instruction located for example at n and then to $n+1$ $n+2$ and so on Exceptions are provided for by the branching instructions described in Table 1

Coding of instructions Instructions are coded in instruction words with the operation (Op) code occupying digit positions 2 and 3 and the address part positions 4 5 and 6 The sign position is 1 and holds a + for all instructions but need not be explicitly written in SC coding

1	2	3	4	5	6	(digital position)
+	X	X	X	X	X	(instruction word format)
	Op code		Address part			

Table 1 contains the instruction list for SC together with explanations

As an example a coded program for SC for evaluating

$$\frac{a_1x_1 + a_2x_2 + \quad + a_{10}x_{10}}{\pi}$$

appears in Table 2 To begin with one (arbitrarily) allocates storage locations for the data as follows a_1 to a_{10} in locations 021 to 030 x_1 to x_{10} in 031 to 040 and π in 050 Also a location is needed in which to accumulate the partial sums as they are formed For this purpose 042 which will also hold the final answer is provided It is assumed that 042 initially contains +00000 The instructions are located in 100 and following addresses In the example the machine language instructions consist only of the instruction words the locutions load X and so on and the remarks are inessential aids to the programmer and any reader of the program Table 2 should be studied using Table 1

Loop programming Loop programming uses a computational sequence over and over again For example the program of Table 2 can easily be shortened by letting the instruction sequence load X multiply add store written once do duty for all pairs a x The resulting program which employs arithmetic operations on instructions so as to modify the address part of 100 and 101 (a procedure called address modification) contains a loop computation will proceed over and over again through the same load X multiply add store sequence until $a_{10}x_{10}$ has been calculated at which time the computation will exit from the loop and proceed to the division An understanding of this procedure is essential to the appreciation of the powerful use and operational character of a digital computer

Loop programming requires the use of a simulated counter which tallies the progress of the index i as (in $a\,x$) a_1x_1 a_2x and so on are successively computed it also requires a test to detect whether the counter has reached its upper limit The test hinges on an exit constant which in the example is +00009 In the following modified program (Table 3) 042 continues to be used as a temporary storage for partial sums 043 is the counter

Table 2. Computation of $\dfrac{\sum_{=1}^{10} a\,x}{\pi}$

Location of instruction		Op code	Address part	Remarks
100	Load X	03	031	Multiplier x_1 placed in X register
101	Mult	04	021	Product $a\,x_1$ formed in Acc
102	Add	01	042	$a\,x_1$ added to partial sum (initially +00000) and stored in 042
103	Store	06	042	
104	Load X	03	032	First four steps repeated for $a\,x_2$ which at 106 is added to a_1x_1 then stored in 042
105	Mult	04	022	
106	Add	01	042	
107	Store	06	042	
108	Load X	03	033	Similarly a_3x_3 computed added to $a\,x_1 + a_2x_2$ and stored in 042
109	Mult	04	023	
110	Add	01	042	
111	Store	06	042	
136	Load X	03	040	$a_{10}x_{10}$ computed added to $\sum^{9} a\,x$ the result is divided by π and stored in 042
137	Mult	04	030	
138	Add	01	042	
139	Divide	05	050	
140	Store X	07	042	

initially holding +00000 and the exit constant +00009 is in location 049

Ignoring for the moment steps 093–099 the instructions located at 100–103 on the first run through the loop do just what the corresponding quadruplet of the program of Table 2 did At 104–105 the counter contents are subtracted from the exit +00009 which as indicated at the bottom of the table is stored at 049 To begin with +00009 − (+00000) = +00009 thus the branch instruction takes the computation out of normal sequence to 111 At 111 113 the constant +00001 (stored at 045) is added to the counter (043)

Next addition is performed on the instructions 100 and 101 (steps 114–119) Because a_2 must now be multiplied by x the address part of 100 and 101 must be changed to 032 and 022 respectively This is accomplished by adding +00001 to the instructions located at 100 and 101 At 120 the computation unconditionally branches to 100 to do the arithmetic for a_2 and x_2

The entire process of computing $a\,x$ testing the counter incrementing the counter and increasing the index i by address modification is continued until the counter is +00009 At this point $a_{10}x_{10}$ has been computed and the result of subtraction at 105 is +00000 so the computation goes in normal sequence to 107 where the evaluation is completed 110 effectively halts computation

Steps 093–099 are necessary to set the counter and temporary storage to +00000 and to initialize instructions 100 and 101 each time the program is used This procedure is absolutely
cause in general programs such as the
either are worked through for many
or repetitiously occur within a

problems The annotated remarks in Table 3 should be considered carefully

Table 3 Loop program

Location of instruction		Op code	Address part	Remarks
093	Clear add	00	046	Counter and temporary storage for partial sums set to +00000
094	Store	06	042	
095	Store	06	043	
096	Clear add	00	047	Address part of 100 and 101 initialized to 031 and 021 respectively
097	Store	06	100	
098	Clear add	00	048	
099	Store	06	101	
				Computation of $a_{k+1}x_{k+1}$ and addition to previously computed $\sum_{i=1}^{k} a_i x$
100	Load X	03	(031)*	
101	Mult	04	(021)*	
102	Add	01	042	
103	Store	06	042	
104	Clear add	00	049	Test contents of counter If Acc >0 branch to 111 if not, continue with 107
105	Sub	02	043	
106	Branch >0	09	111	
107	Clear add	00	042	Divide by π answer in 042
108	Div	05	050	
109	Store X	07	042	
110	Un branch	08	110	Loops at 110 (stops computation)
111	Clear add	00	043	Counter increased by +00001
112	Add	01	045	
113	Store	06	043	
114	Clear add	00	100	Address part of 100 increased by +00001
115	Add	01	045	
116	Store	06	100	
117	Clear add	00	101	Address part of 101 increased by +00001
118	Add	01	045	
119	Store	06	101	
120	Un branch	08	100	Return to compute $a_{k+2}x_{k+2}$
Constants				
045		+00	001	
046		+00	000	
047		+03	031	
048		+04	021	Counter at 043
049		+00	009	Temporary storage at 042

* Parentheses denote address subject to modification or variable address

In the foregoing no attention has been paid to the problems of fixing the decimal point the use of floating point representation overflow minimum access coding or to the problem of getting data in and out of SC In addition many machines are equipped with index registers which enable an automatic procedure for modifying instructions For these matters the references should be consulted For discussion of the actual computer operation *see* DIGITAL COMPUTER

SC has a complete instruction code (except for input and output) in the sense that a problem computable on any digital computer can be programmed (to within storage capacity) on SC Nevertheless practical computers usually have many additional instructions such as indexing shifting and other logical operations for convenience

SC is a single address machine in that instructions make reference to but one address Two-, three, or even four address machines are possible For example, a three address machine might have such an instruction as "Add *a* to *b* and store result in *c*" Such a scheme requires longer words more elaborate arithmetical and control means within the computer and is more wasteful of memory space although it is perhaps easier to program

Subroutine A subroutine is a program to which a main program branches and from which the main program takes over control when the routine is completed (Sometimes such programs are called closed subroutines) Frequently used arithmetic operations such as root extraction for which no machine instruction is usually provided may be permanently programmed and branched into and out of other programs as required An example which incidentally shows how useful basic operations lacking in SC may be programmed is the subroutine (Table 4) for branching on zero It is desired to simulate an instruction that would read "Branch 0 YYXXX" where YY would be the Op code and XXX the address part as usual and where the intention is "if Acc is +00000 go to XXX, otherwise continue in normal sequence"

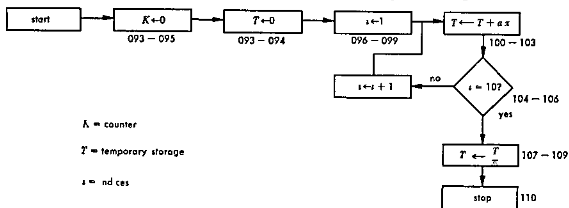
It is necessary to provide means whereby the subroutine can find its way back to the main program In the example of Table 4 the last three instructions of the main program together with the first four of the branch 0 routine constitute a linkage for return Essentially the main program leaves its own next to last instruction in the accumulator and the subroutine adds an appropriate constant which in effect constructs the exit instruction (an unconditional branch) from which the subroutine will itself branch back to the main program The example although it is not easy to follow is better than extended exposition It is assumed that the subroutine begins at location 020, the main program leaves off at 103 and is to resume at 104 if the simulated branch 0 test fails or at 110 if it succeeds

The linkage is independent of the main program's being at 102 or any other location when ready for the subroutine The exit instructions at 028 and 029 contain address parts in parentheses to indicate variability Location 200 is used of course in all applications of the routine

Programming aids Flow charts are used in programming computations which are too complicated to be planned easily during actual coding of instructions A flow chart for the program of Table 3 is given in the illustration All operations are symbolized by an arrow indicating replacement of contents of a storage location (or other register) by the datum shown with or without an indicated arithmetical operation Branching is represented by a diamond shaped decision box

To aid in following the chart the corresponding instruction locations of the coded program of Table 3 are shown

Programming aids of various kinds have been developed for reducing errors inherent in coding and address assignments, and for relegating the task of machine language coding as in the foregoing examples to the computer itself



Flow chart of program in Table 3

In relative programming storage regions or blocks are allocated for constants data instruc tions subroutines and the like and the program mer assigns relative to a region addresses which are not the final storage locations or absolute ad dresses for the running program An assembly pro gram is then used to assign absolute addresses and turn out a fully coded program A symbolic assem bly program is similar except that the addresses may be alphabetic symbols of a highly mnemonic character for example PI for the address of π

Any scheme which enables the programmer to translate from a flow chart into machine language by means of a simplified and more humanly man ageable language called a pseudo code is called an automatic program Pseudo instructions are either executed by an interpretive routine which inter prets the instructions and acts on them by means of special subroutines one by one or translated and

Table 4 Subrout ne for branch 0

Locat on of instruction		Op code	Address part	Remarks
099 (γ, α)				
100	(Acc holds number to be tested in branch)			
101	Store	06	200	Number to be tested for + stored
102	Clear add	00	102	102 places itself in accumulator
103	Un branch	08	020	Branch to subrou t ne
020	Add	01	040	Construct ex t to main program if non 0 (00102 + 08002 = 08104)
021	Store	06	028	
022	Add	01	041	Construct ex t to main program if 0 (08104 + 00006 = 08110)
023	Store	06	029	
024 (β)	Clear add	00	200	Branch 0 subroutine proper
025	Branch >0	09	028	
026	Add	01	042	
027	Branch >0	09	029	
028	Un branch	08	(104)*	Exits
029	Un branch	08	(110)*	
			Constants	
α = Ma n program				
β = Subrout ne		040 08002		041 00006
γ = Linkage		042 00001		

* Parentheses denote address subject to modif cation or var able address

compiled by a program that produces a complete machine language program [R J N]

Bibliography R V Andree *Programming the IBM 650 Magnetic Drum Computer and Data Processing System* 1958 D D McCraken *Digital Computer Programming* 1957 M V Wilkes D J Wheeler and S Gill *The Preparation of Programs for an Electronic Digital Computer* 2d ed 1957

Digital voltmeter

A voltmeter with electronic counting circuits used to give a digital indication of applied voltage A common circuit arrangement employs a ramp func tion generator in which a voltage periodically starts from zero and increases at a uniform and accurately controlled rate An electronic counter counts the cycles of a precision oscillator between the starting instant and the instant when the ramp voltage equals the applied voltage the count then being proportional to the voltage Some instru ments use an in line presentation of illuminated numerals while others employ a column of 10 lights at each position one light of each decade being turned on to give the corresponding digit Most in struments have automatic positioning of the deci mal point and indication of positive or negative po larity *See* NUMBER INDICATOR TUBE VOLTAGE MEASUREMENT [W N T]

Digitalis

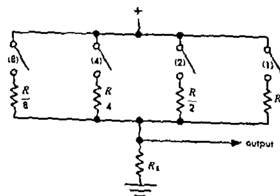
A genus of the figwort family (Scrophulariaceae) ranging from the Canary Islands to central Asia *Digitalis purpurea* or foxglove a native of western Europe is the source of the important drug digi talis much used in the treatment of heart disorders Fresh mature leaves are carefully selected quickly dried and stored in airtight containers The active ingredient of digitalis is the glucoside digitalin This slows and regulates the heartbeat improving the tone and rhythm and making the contractions more effective The plant is also prized as a herba ceous perennial *See* TUBIFLORALES [P D S]

Digital to analog converter

A device for converting information in the form of discrete (usually binary) signals to analog form For discussion of analog and digital devi COMPUTER

Digital to analog converters are used most frequently to present the results of digital computation either for graphical display or for control of devices that operate with continuously varying quantities

Many conversion devices consist of voltage divider networks The figure shows a simple converter for transforming from a digital 8 4 2 1 code to an analog voltage output Each of the four switches represents the binary 1 when closed and 0 when open The decimal number 5 for example is represented by closing the (4) and the (1) switch A current proportional to the sum of the weights (4 + 1) flows through the summing resistor R_L at the bottom of the figure The voltage appearing at the output is proportional to the value of the number and is therefore an analog representation of that number This kind of system will work only with weighted codes or the straight binary system For a discussion of binary codes *see* DIGITAL COMPUTER In the example the four digital input of the switches (which may be relays gates or other binary devices) may be outputs of flip flops relays or other digital storage devices



Digital-to-analog voltage network converter

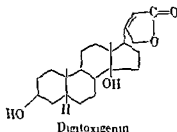
In order to change a position or rotate a shaft in control applications an analog converter is made part of a feedback system The position is digitally *encoded* and fed into an ordinary digital subtractor where the difference *between the digital* control signal and the encoded posit... tal differe... plified an... loop arra... ...of the position encoder becomes equal to the digital number used to establish the position [R J N]

ita

Or

Digitoxigenin

A steroid isolated from the seeds or leaves of *Digitalis purpurea* It exists as a glycosidic derivative in the plant In this combined form of glycoside it is highly toxic but in proper dosage *it is of value* in the control of arteriosclerotic and hypertensive

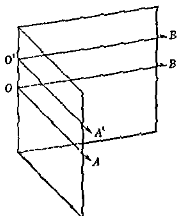


Digitoxigenin

heart disease especially when auricular fibrillation is present *See* STEROID, STEROL [R I D]

Dihedron

A geometric figure formed by two half planes that are bounded by the same straight line This line is called the edge of the dihedron Planes perpendicular to the edge cut the dihedron in equal plane



Dihedron and dihedral angle

angles *AOB* and *A'O'B'* whose common measure is called the dihedral angle of the dihedron Two dihedral angles are supplementary if their sum is 180° complementary if their sum is 90° When two planes intersect they form four dihedrons of which ... dihedral ... ruent and [J S F]

Dill

A small annual or biennial herb *Anethum graveolens*, of high repute among ancient peoples and now cultivated in Europe, India, and the United States In the United States, dill is used mainly as a flavoring for pickles In France India and other countries it is used in soups and sauces stews and other dishes Both the fruits and oil are used in medicine *See* SPICE AND FLAVORING, UMBELLALES [P D S]

Dimensional analysis

A technique that involves the study of dimensions of physical quantities Dimensional analysis is used primarily as a tool for obtaining information about physical systems too complicated for full mathematical solutions to be feasible It enables one to predict the behavior of large systems from a study of small scale models It affords a conven

ient means of checking mathematical equations Finally dimensional formulas provide a useful cataloging system for physical quantities

Theory *All* the commonly used systems of units in physical science have the property that the number representing the magnitude of any quantity (other than purely numerical ratios) varies inversely with the size of the unit chosen Thus if the length of a given piece of land is 300 ft its length in yards is 100 The ratio of the magnitude of 1 yard to the magnitude of 1 foot is the same as that of any length in feet to the same length in yards that is 3 The ratio of two different lengths measured in yards is the same as the ratio of the same two lengths measured in feet inches miles or any other length units This universal property of unit systems often known as *the absolute significance of relative magnitude* determines the structure of all dimensional formulas *See* UNITS SYSTEMS OF

In defining a system of units for a branch of science such as mechanics or electricity certain quantities are chosen as fundamental and others as secondary or derived The choice of the fundamental units is always arbitrary and is usually made on the basis of convenience in maintaining standards In mechanics the fundamental units most often chosen are mass length and time Standards of mass (the standard kilogram) and of length (the *standard meter*) are readily manufactured and preserved while the rotation of the earth gives a sufficiently reproducible standard of time Secondary quantities such as velocity force and momentum are obtained from the primary set of quantities according to a definite set of rules

Assume that there are three primary or fundamental quantities α β and γ (the following discussion however is not limited to there being exactly three fundamental quantities) Consider a particular secondary quantity expressed in terms of the primaries as $F(\alpha\,\beta\,\gamma)$ where F represents some mathematical function For example if $\alpha =$ mass $\beta =$ length and $\gamma =$ time the derived quantity velocity would be $F(\alpha\,\beta\,\gamma) = \beta/\gamma$

Now if it is assumed that the sizes of the units measuring α β and γ are changed in the proportions $1/x$ $1/y$ $1/z$ respectively then the numbers measuring the primary quantities become $x\alpha$ $y\beta$ $z\gamma$ and the secondary quantity in question becomes $F(x\alpha\ \ y\beta\ \ z\gamma)$ Merely changing the sizes of the units must not change the rule for obtaining a particular secondary quantity

Consider two separate values of the secondary quantity $F(\alpha_1\,\beta_1\,\gamma_1)$ and $F(\alpha_2\,\beta_2\,\gamma_2)$ Then according to the principle of the absolute significance of relative magnitude

$$\frac{F(x\alpha_1\,y\beta_1\,z\gamma_1)}{F(x\alpha_2\,y\beta_2\,z\gamma_2)} = \frac{F(\alpha_1\,\beta_1\,\gamma_1)}{F(\alpha_2\,\beta_2\,\gamma_2)}$$

or $$F(x\alpha_1\,y\beta_1\,z\gamma_1) = F(x\alpha_2\,y\beta_2\,z\gamma_2)\,\frac{F(\alpha_1\,\beta_1}{F(\alpha_2\,\beta}$$

Differentiating partially with respect to x holding y and z constant gives

$$\alpha_1 F'(x\alpha_1\,y\beta_1\,z\gamma_1) = \alpha_2 F'(x\alpha_2\,y\beta_2\,z\gamma_2)\,\frac{F(\alpha_1\,\beta_1\,\gamma_1)}{F(\alpha_2\,\beta_2\,\gamma_2)}$$

where F' represents the total derivative of the function F with respect to its first argument

Next set the coefficients x y $z = 1$ This gives

$$\frac{\alpha_1 F'(\alpha_1,\beta_1\,\gamma_1)}{F(\alpha_1\,\beta_1\,\gamma_1)} = \frac{\alpha_2 F'(\alpha_2\,\beta_2\,\gamma_2)}{F(\alpha_2\,\beta_2\,\gamma_2)}$$

This relation must hold for all values of the arguments α β and γ and hence is equal to a constant The subscripts can now be dropped giving

$$\frac{\alpha}{F}\frac{dF}{d\alpha} = \text{constant}$$

The general solution of this differential equation is $F = C_1\alpha^a$ where a is a constant and C_1 is in general a function of β and γ

The above analysis can now be repeated for the parameters β and γ leading to the results

$$F = C_2(\alpha\,\gamma)\beta^b$$
$$F = C_3(\alpha,\beta)\gamma^c$$

These solutions are consistent only if $F = C\,\alpha^a\beta^b\gamma^c$ where C a b and c are constants Thus every secondary quantity which satisfies the condition of *the absolute significance of relative magnitude* is expressible as a product of powers of the primary quantities Such an expression is known as the dimensional formula of the secondary quantity There is no requirement that the exponents a b c be integral

Examples of dimensional formulas Table 1 gives the dimensional formulas of a number of mechanical quantities in terms of mass M length L and time T

In order to extend this list to include the dimensional formulas of quantities from other branches

Table 1 Dimensional formulas of common quantities

Quantity	Definition	Dimensional formula
Mass	Fundamental	M
Length	Fundamental	L
Time	Fundamental	T
Velocity	Distance/time	LT^{-1}
Acceleration	Velocity/time	LT^{-2}
Force	Mass × acceleration	MLT^{-2}
Momentum	Mass × velocity	MLT^{-1}
Energy	Force × distance	ML^2T^{-2}
Angle	Arc/radius	0
Angular velocity	Angle/time	T^{-1}
Angular acceleration	Angular velocity/time	T^{-2}
Torque	Force × lever arm	ML^2T^{-2}
Angular momentum	Momentum × lever arm	ML^2T^{-1}
Moment of inertia	Mass × radius squared	ML^2
Area	Length squared	L^2
Volume	Length cubed	L^3
Density	Mass/volume	ML^{-3}
Pressure	Force/area	[illegible]
Action	Energy × time	[illegible]
Viscosity	Force per unit area per unit velocity gradient	[illegible]

of physics such as electricity and magnetism one may do either of the following

1 Obtain the dimensional formulas in terms of a particular unit system without introducing any new fundamental quantities Thus if one uses Coulomb s law in the centimeter gram second electrostatic system in empty space (dielectric constant = 1) one obtains

$$\text{Force} = \frac{(\text{charge})^2}{(\text{distance})^2}$$

or Charge = distance × (square root of force)

Thus the dimensions of charge are $M^{1/2}L^{3/2}T^{-1}$ The dimensional formulas of other electrical quantities in the electrostatic system follow directly from the definitions

$$\text{Potential} = \text{energy/charge} = M^{1/2}L^{1/2}T^{-1}$$
$$\text{Current} = \text{charge/time} = M^{1/2}L^{3/2}T^{-2}$$
$$\text{Electric field} = \text{force/charge} = M^{1/2}L^{-1/2}T^{-1}$$

and so forth See ELECTRICAL UNITS

2 Introduce an additional fundamental quantity to take account of the fact that electricity and magnetism encompass phenomena not treated in mechanics All electrical quantities can be defined in terms of M L T and one other without resorting to fractional exponents and without the artificial assumption of unit (and dimensionless) *dielectric constant* as in the electrostatic system The usual choice for the fourth fundamental quantity is charge Q even though charge is not a preservable electrical standard Then

$$\text{Potential} = \text{energy/charge} = ML^2T^{-2}Q^{-1}$$
$$\text{Current} = \text{charge/time} = QT^{-1}$$

and so forth

It must be realized that the choice of *fundamental quantities* is entirely arbitrary For example a system of units for mechanics has been proposed *in which the velocity of light is taken as a dimensionless quantity* equal to unity in free space All velocities are then dimensionless All mechanical *quantities* can then be specified in terms of just two fundamental quantities mass and time This is analogous to the *reduction of the* number of fundamental quantities needed for electricity from four to three by taking the dielectric constant (permittivity) of empty space as *unity*

Furthermore one could increase the number of fundamental quantities in *mechanics* from three to four by adding force F to the list and rewriting Newton s second law as $F = Kma$ where K is a constant of dimensions $M^{-1}L^{-1}T^{2}F$ One could then define a system of units for which K was not numerically equal to 1

In the past there has been considerable controversy as to the absolute significance *if* any of dimensional formulas The significant fact is that dimensional formulas always *consist of* products of powers of the fundamental *quantities*

Applications The important uses of the technique of dimensional analysis are *considered* in the following sections

Checking of equations It is intuitively obvious that only terms whose dimensions are the same can be equated The equation 10 kilograms = 10 meters per second for example makes no sense A necessary condition for the correctness of any equation is that the two sides have the same dimensions This is often a help in the verification of complicated analytic expressions Of course an equation can be correct dimensionally and still be wrong by a purely numerical factor

A corollary of this is that one can add or subtract only quantities which have the same dimensions (except for the trivial case which arises when two different equations are added together) Furthermore the arguments of trigonometric and logarithmic functions must be dimensionless otherwise their power series expansions would involve sums of terms with different dimensions There is no restriction on the multiplication and division of terms whose dimensions are different

Derivation of equations—the π theorem The application of dimensional analysis to the derivation of unknown relations depends upon the concept of completeness of equations An expression which remains formally true no matter how the sizes of the fundamental units are changed is said to be complete If changing the units makes the expression wrong it is incomplete For a body starting from rest and falling freely under gravity $s = 16t^2$ is a correct expression only so long as the distance fallen s is measured in feet and the time t in seconds If s is in meters and t in minutes the equation is wrong Thus $s = 16t^2$ is an incomplete equation The constant in the equation depends upon the units chosen To make the expression complete the numerical factor must be replaced by a dimensional constant the acceleration of gravity g Then $s = \frac{1}{2}gt^2$ is valid no matter how the units of length and time are changed since the numerical value of g can be changed accordingly

Assume a group of n physical quantities x_1 x_2 … x_n for which there exists one and only one complete mathematical expression connecting them namely $\phi(x_1,x_2, \ldots ,x_n) = 0$ Some of the quantities x_1 x_2 … x_n may be dimensional constants Assume further that the dimensional formulas of the n quantities are expressed in terms of m fundamental quantities α β γ … Then it will always be found that this single relation ϕ can be expressed in terms of some arbitrary function F of $n-m$ independent dimensionless products π_1 π_2 … π_{n-m} made up from among the n variables

$$F(\pi_1, \pi_2, \ldots, \pi_{n-m}) = 0$$

This is known as the π theorem It was first rigorously proved by E Buckingham The proof is straightforward but long The only restriction on the πs is that they be independent (no one expressible as products of powers of the others) The number m of fundamental quantities chosen in a particular case is immaterial so long as they also are dimensionally independent Increasing m by 1 al

ways increases the number of dimensional constants (and hence n) by 1 leaving $n - m$ the same

The main usefulness of the π theorem is in the deduction of the form of unknown relations The successful application of the theorem to a particular problem requires a certain amount of shrewd guesswork as to which variables x_1 x_2 x_n are significant and which not If $\phi(x_1, x_2 \quad , x_n)$ is not known one can often still deduce the structure of $F(\pi_1 \pi_2 \quad \pi_{n-m}) = 0$ and so obtain useful information about the system in question

An example is the swinging of a simple pendulum Assume that the analytic expression for its period of vibration is unknown Choose mass length and time as the fundamental quantities Thus $m = 3$ One must make a list of all the parameters pertaining to the pendulum which might be significant If the list is incomplete no useful information will be obtained If too many quantities are included the derived information is less specific than it might otherwise be The quantities given in Table 2 would appear to be adequate

Table 2

Quantity	Symbol	Dimensional formula
Mass of bob	m	M
Length of string	l	L
Acceleration of gravity	g	LT^{-2}
Period of swing	τ	T
Angular amplitude	θ	0

The microscopic properties of the bob and string are not considered and air resistance and the mass of the string are likewise neglected These would be expected to have only a small effect on the period Thus $n = 5$ $n - m = 2$ and therefore one expects to find two independent dimensionless products These can always be found by trial and error There may be more than one possible set of independent πs In this case one π is simply θ the angular amplitude Another is $l/\tau^2 g$ Since no other of the n variables contains M in its dimensional formula the mass of the bob m cannot occur in any dimensionless product The π theorem gives $F(\theta\, l/\tau^2 g) = 0$ Therefore the period of vibration does not depend upon the mass of the bob

Because τ appears in only one of the dimensionless products this expression can be explicitly solved for τ to give $\tau = G(\theta)\sqrt{l/g}$ where $G(\theta)$ is an arbitrary function of θ Now make the further assumption that θ is small enough to be neglected Then $n = 4$ $n - m = 1$ and $F(l/\tau^2 g) = 0$ Thus $l/\tau^2 g =$ constant or $\tau \propto \sqrt{l/g}$ The π theorem thus leads to the conclusion that τ varies directly as the square root of the length of the pendulum and inversely as the square root of the acceleration of gravity The magnitude of the dimensionless constant (actually it is 2π) cannot be obtained from dimensional analysis

In more complicated cases where a direct solution is not feasible this method can give information on how certain variables enter a particular problem even when $F(\pi_1 \pi_2 \quad \pi_{n-m}) = 0$ cannot be explicitly solved for one variable The procedure is particularly useful in hydraulics and aeronautical engineering where detailed solutions are often extremely complicated

Model theory A further application of dimensional analysis is in model design Often the behavior of large complex systems can be deduced from studies of small scale models at a great saving in cost In the model each parameter is reduced in the same proportion relative to its value in the original system

Once again the case of the simple pendulum is a good example It was found from the π theorem that $F(\theta\, l/\tau^2 g) = 0$ If the magnitudes of θ l τ and g are now changed in such a way that neither argument of F is changed in numerical value the system will behave exactly as the original system and is said to be physically similar Evidently θ cannot be changed without altering any of the arguments of F but l g and τ can be varied Suppose that it was desired to build a very large and expensive pendulum which was to swing with finite amplitude One could build a small model of say 1/100 the length and time its swing for an amplitude equal to that for the desired pendulum The acceleration of gravity g would be the same for the model as for the large pendulum The period for the model would then be just 1/10 that for the large pendulum Thus the period of the large pendulum could be deduced before the pendulum was ever built In practice one would never bother with the π theorem in cases as simple as this where a full analytic solution is possible In many situations where such a solution is not feasible models are built and extensively studied before the full scale device is constructed This technique is standard in wind tunnel studies of aircraft design *See* DYNAMIC SIMILARITY

Cataloging of physical quantities Dimensional formulas provide a convenient shorthand notation for representing the definitions of secondary quantities These definitions depend upon the choice of primary quantities The π theorem is applicable no matter what the choice of primary quantities is

Changing units Dimensional formulas are helpful in changing units from one system to another For example the acceleration of gravity in the centimeter gram second system of units is 980 cm/sec² The dimensional formula for acceleration is LT^{-2} To find the magnitude of g in mi/hr² one would proceed as follows

$$980\,\frac{\text{cm}}{\text{sec}^2} \times \frac{\text{(conversion factor for length)}}{\text{(conversion factor for time)}^2}$$

$$= 980\,\frac{\text{cm}}{\text{sec}^2} \times \frac{\dfrac{1}{1\,609 \times 10^5}\,\dfrac{\text{miles}}{\text{cm}}}{(1/3600)^2(\text{hr}^2/\text{sec}^2)}$$

$$= 7.89 \times 10^5\ \text{mi/hr}^2$$

In the past the subject of dimensions has been quite controversial For years unsuccessful attempts were made to find ultimate rational quantities in terms of which to express all dimensional formulas It is now universally agreed that there is no one absolute set of dimensional formulas Some systems are more symmetrical than others and for this reason are perhaps preferable The representation of electrical quantities in terms of M L and T alone through the electrostatic form of Coulomb s law leads to somewhat awkward fractional exponents but nevertheless is just as correct as a representation in which charge is used as a fourth fundamental unit

A highly symmetrical pattern results if energy linear displacement and linear momentum are chosen as the fundamental quantities in mechanics In electricity one can use energy charge and magnetic flux The corresponding quantities for a vibrating mass on a spring and the analogous alternating current circuit with inductance and capacitance have similar dimensional formulas In this analogy energy is invariant charge corresponds to displacement and magnetic flux corresponds to linear momentum This correspondence is not displayed in conventional dimensional formulas *See* DYNAMICAL ANALOGIES [J W ST]

Bibliography P W Bridgman *Dimensional Analysis* 1931 W J Duncan *Physical Similarity and Dimensional Analysis* 1953 A W Porter *The Method of Dimensions* 1946

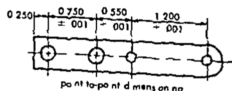
Dimensioning

[illegible]

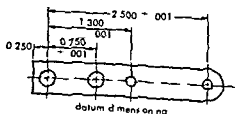
[illegible]ion of a detail such as a hole or boss Units for machine part dimensions are inches or when over 72 in feet and inches *Dimensions* specify the size and shape of the part as required by the designer and aid the workman in constructing the part

Two plans of dimensioning are used as illustrated One is called *point to point* or chain dimensioning each length is dimensioned from the end of the preceding one These dimensions are usually taken directly from the designer s sketch however tolerances (*see* TOLERANCE) are cumulative and rather than averaging out, usually add or subtract so that intermediate clearances may be disturbed In the other plan called datum dimensioning or the reference line method all dimension lines in each direction extend from a datum or reference line or plane which is usually the first machined surface Here tolerances are not cumulative and a mating part will be within the specified limits *See* DRAFTING [P H B]

Bibliography American Society of Mechanical Engineers *Drafting Standards Manual Section 5 Dimensions and Notes* ASA Y14 5–1957



point-to-point dimensioning



datum dimensioning

Two standard methods for marking dimensions on mechanical drawings

Dimensions (mechanics)

Length mass time or combinations of these quantities serving as an indication of the nature of a physical quantity Quantities with the same dimensions can be expressed in the same units For example although speed can be expressed in various units such as miles/hour feet/second meters sec and so on all these speed units involve the ratio of a length unit to a time unit hence the dimensions of speed are the ratio of length L to time T usually stated as LT^{-1} The dimensions of all mechanical quantities can be expressed in terms of L T and mass M The validity of algebraic equations involving physical quantities can be tested by a process called dimensional analysis the terms on the two sides of any valid equation must have the same dimensions *See* DIMENSIONAL ANALYSIS *see also* UNITS SYSTEMS OF [D WI]

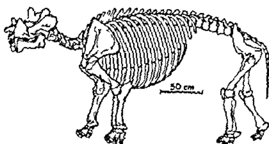
Dinocerata

An extinct order of large herbivorous mammals the uintatheres from early Cenozoic deposits of North America and northeastern Asia Members of this group have semigraviportal limbs with hoofed 5 toed feet The dentition is somewhat reduced in all forms later forms losing all the upper incisors and even in one case (*Gobiatherium* late Eocene Mongolia) the upper canine The upper molars and premolars are V shaped The lower molars and premolars possess V shaped crests followed by a low shelf A saberlike canine tooth and protective lower jaw flange are present in all forms except the aberrant *Gobiatherium* which must have relied on other means for defense Horns are absent on the most primitive forms but by early Eocene time North American genera had begun to develop them Middle and late Eocene uintatheres in North America developed an imposing array of six horns one pair on the tips of the nasal bones another above the root of the saberlike canine tooth and a third pair above the ear region

The Dinocerata may be divided into three families Prodinoceratidae Uintatheriidae and Gobiatheriidae Prodinoceratidae has been proposed to include two Mongolian genera from the late Paleo-

cene and early Eocene an American genus from deposits of the same ages and tentatively a second American genus from the late Paleocene *Mongolotherium* is the best known of these and demonstrates that the most primitive uintatheres possessed a carnivorelike body of moderate size

The Uintatheriidae in the restricted sense or horned forms includes all other American members of the order Uintatheres increased to rhinoceros size in the Eocene were especially common in the middle Eocene and died out before the Oligocene began



Skeleton of *Uintatherium* a middle Eocene member of the Dinocerata (After Flerov)

Gobiatheriidae an aberrant side branch from the early Uintatheriidae is sufficiently distinct to merit family rank This family known from one late Eocene Mongolian genus is characterized by extreme reduction of the anterior dentition and lack of horns The skull is remarkably low and flat

The Dinocerata left no descendants and are believed to have arisen from the arctocyonid creodont carnivores but from a different subfamily than that which gave rise to the order Pantodonta *See* CARNIVORA FOSSILS PANTODONTA [M C MC]

Dinoflagellida

An order of the class Phytomastigophorea also known as the Dinoflagellata *Noctiluca scintillans* one of the largest species may measure 1 5 mm whereas the flagellate stage of some species may be as small as 10 μ Although primarily marine some dinoflagellates occur in fresh water Some possess brown chromatophores some are variously colored and others are colorless Masking pigments are frequent being cytoplasmic or in chromatophores All types of nutrition exist and some species are parasitic Two flagella emerge laterally from a longitudinal depression or sulcus An encircling flagellum in a groove or girdle divides the body into epicone and hypocone the other extending backward propels the organism forward The nucleus is very large

Dinoflagellate species often of bizarre form have fixed shapes determined by thick covering plates except in Gymnodinia which are naked or have thin pellicles Chain formation ameboid and palmella aggregates occur *Ceratium hirundinella* (Fig 1) blooms in fresh water causing tastes and

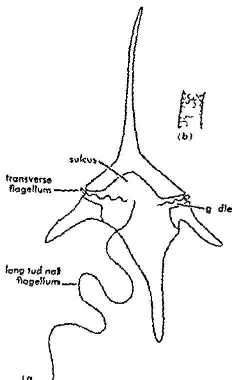


Fig 1 (a) *Ceratium hirundinella* Size 95–700 μ (b) Pitted surface

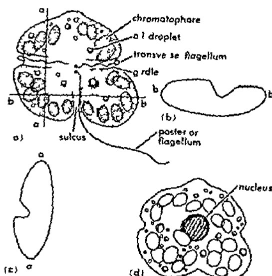


Fig 2 (a) *Gymnodinium breve* Varies from 28 to 45 μ (b) Section to show sulcus (c) Section to show girdle (d) *Gymnodinium breve* fixed to show the central nucleus

odors Marine blooms of *Gymnodinium breve* (Fig 2) produce the fish killing red tides which occur along the Atlantic Coast of the southern United States *See* PHYTOMASTIGOPHOREA [J B L]

Dinornithiformes

An order of birds containing 2 families with 23 species divided among 5 genera the extinct moas of New Zealand Fossil moas are known from Upper Miocene and Lower Pliocene deposits and the last surviving species are thought to have been exterminated by pre Maori Polynesian colonists of New Zealand A supposed Pleistocene species from Australia is now identified with the emus Remains of about 20 species have been found ranging from turkey sized birds to giants over 10 ft tall All had strong legs with 4-toed feet and wings that were vestigial or absent Moas had long necks and small heads their sternum lacked a keel and both scapula and coracoid were fused into a slender curved rod The feathers resemble those of the emu which had a large aftershaft and loose webs and lacked terminal barbicels Some were brown with a black subterminal band and white tip others had yellow center stripes or were white Moas are known from bones fragments of dried muscle and ligament feathers eggshells and gizzard stones Remains have been found at ancient human campsites Their closest relatives appear to be the living kiwis (Apterygiformes) also of New Zealand See AVES AVES FOSSILS NEORNITHES [A W K C P]

Dinosaur

The name meaning terrible lizard given to the fossil bones of ancient reptiles many times the size of the largest living crocodiles The varied land dwelling reptiles of the Mesozoic Era to which this term is properly applied are now placed in the subclass Archosauria as two separate orders the Saurischia and Ornithischia distinguished by the arrangement of the hip bones whether reptilian or birdlike (Fig 1) The term dinosaur long ago taken into the vernacular has largely disappeared from formal classifications Among the dinosaurs were the largest land dwelling animals that the world has known but not all were large indeed one known specimen is no larger than a chicken See ARCHOSAURIA

The essential features which distinguish the dinosaurian orders from more primitive archosaurs and other reptiles are the limbs and related portions of the skeleton Early dinosaurs were bipedal with long rear legs and a heavy tail which balanced the fore part of the body on the hips as a fulcrum A firm union between the hip bones and back bone involving an enlargement of the sacrum to include more than the usual two vertebrae resulted from this habit Some later members of each order reverted to quadrupedal gait but retained indications of their bipedal ancestry in the structure of the pelvis and sacrum and with few exceptions in their relatively short forelimbs

Saurischia Saurischian dinosaurs include both carnivorous and herbivorous types Bipedal saurischians had birdlike feet with three toes directed ... a heel ... nded in ... ropriate name Theropoda which means beast footed Most if not all theropods were carnivorous and their jaws were rimmed with sharp pointed compressed knifelike teeth often with serrate edges

Quadrupedal saurischians the largest dinosaurs were strange beasts with elephantine bodies and limbs long slender necks and whiplash tails They constitute the suborder Sauropoda an equally inappropriate name meaning lizard or reptile footed The sauropods were all herbivorous

The earliest adequately known theropod *Coelophysis* from Late Triassic rocks of southwestern North America reached a length of 2 5 m (Fig 2) It had a long narrow pointed skull a rather long neck relatively short forelimbs with three long fingers and much longer strong slender rear limbs suitable for running From such an ancestor developed the larger carnivorous dinosaurs of the Juras

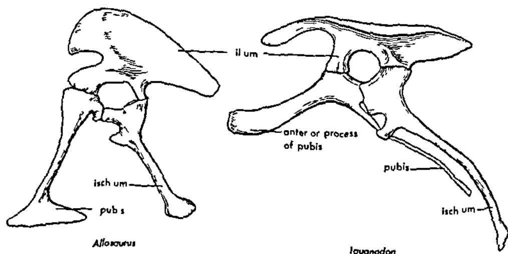
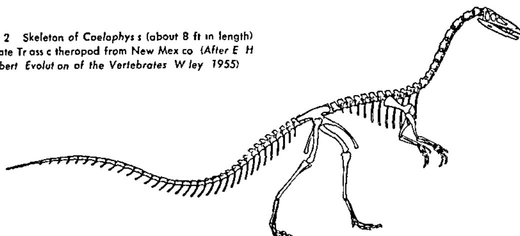


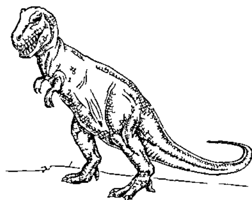
Fig 1 Hip bones of Saurischian (*Allosaurus*) and Ornithischian (*Iguanodon*) dinosaurs (After O C Marsh and L Dollo)



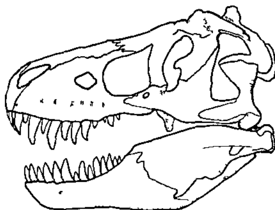
Fg 2 Skeleton of *Coelophys s* (about 8 ft in length) a Late Tr ass c theropod from New Mex co (*After E H Colbert Evolut on of the Vertebrates W ley 1955*)

sic and Cretaceous animals which differed from *Coelophysis* principally in having shorter necks relatively larger and heavier skulls even smaller forelimbs and more massive rear limbs to support the r greater weight *Megalosaurus* from Jurassic rocks of Europe was 3–7 m long and retained five toes in the forefoot *Allosaurus* (Fig 3) a similar animal from the Late Jurassic of North America was nearly 11 m long but had only three fingers *Ceratosaurus* also from North America was about one half the size of *Allosaurus* but had a horn on its nose *Tyrannosaurus* from Late Cretaceous of North America and Mongolia the largest of all theropods had a skull 1 5 m long (Fig 4) a body 16 m from nose to tail and tiny forelimbs bear ng only two claws these were the great flesh eaters of the Mesozoic world Other theropods did not grow so large *Struthiomimus* of the Late Creta ceous retained the elongate neck and fa rly well developed forelimbs of *Coelophysis* but had no teeth in its jaws which probably were covered by a birdlike beak There has been much speculation about whether it fed on plants eggs or small ani mals which t caught with its slender fingers

Sauropod d nosaurs had blunt spoon shaped teeth which tended to be confined to the front of



Fg 3 Restorat on of the b pedal carn vorous saur s ch an *Allosaurus* (about 40 ft long) Late Jurass c of North Amer ca



Fg 4 Skull of *Tyrannosaurus* (about 5 ft long) larg est of the carn vorous theropods Upper Cretaceous of North Amer ca and eastern As a (*After H F Osborn*)

the mouth and apparently served principally for cropping vegetat on The earliest members of this suborder such as *Plateosaurus* have sometimes been classified with the theropods because they re tained at least partially a bipedal posture How ever the forelimbs were longer the first toe of the rear foot was not turned backward as a heel support and the 5 toed hand bore an enormous claw on the first finger

Later sauropods had longer necks the skulls be came relatively small To support their great we ght the limb bones became solid and pillarlike and the forelimb increased its size so that in *Brachiosaurus* the forelimb actually exceeded the rear in length To compensate in part for their great weight the vertebrae developed bone only along lines of mechanical stress and consisted of a series of ridges and hollows Some of the cavities n the centra of sauropods may well have contained air sacs similar to those in the wing bones of b rds

Plateosaurus of the Late Triassic of Germany was 6 m long *Anchisaurus* was a similar but slightly smaller North American contemporary *Diplodocus* (Fig 5) of the Late Jurassic of North Americ

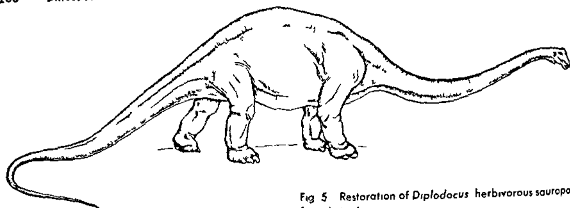
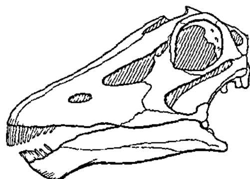


Fig 5 Restoration of *Diplodocus* herbivorous sauropod from Late Jurassic

the greatest body length about 30 m its neck was extremely elongate Its nostrils were far back between the eye openings a position that suggests aquatic habits The skull of *Diplodocus* was solidly built rather long and low with a few slender cropping teeth at the front of the mouth (Fig 6*a*)

The somewhat smaller Jurassic sauropod *Camarasaurus* had a shorter higher more lightly constructed skull (Fig 6*b*) and a larger number of large blunt spoon shaped teeth Its greatly enlarged nostrils were raised above the level of the skull roof *Brachiosaurus* known from Late Jurassic deposits of both North America and Africa and notable for its relatively long forelimbs had a skull somewhat similar to that of *Camarasaurus* as did *Helopus* from the Early Cretaceous of China The Cretaceous sauropod of South America *Antarctosaurus* had a skull constructed more on the pattern of *Diplodocus* The most widely known sauropod of all *Brontosaurus* has never been found with skull attached so its relationships to other genera are uncertain Commonly it has been restored with a head similar to a *Camarasaurus* but the vertebrae and limbs show more resemblance to *Diplodocus*



(a)



(b)

Fig 6 Skulls of the sauropod dinosaurs (a) *Diplodocus* (b) *Camarasaurus* (After O C Marsh and C W Gilmore)

Ornithischia Ornithischian or bird hipped dinosaurs were entirely herbivorous A horny beak was developed at the front of the mouth covering both the premaxillary (toothless except in the primitive ornithopod genus *Hypsilophodon*) and a special median bone the predentary in the lower jaw Further back in the jaws were lancet shaped teeth with sharp denticulate edges The toes ended in rounded or blunt hooves instead of claws To the Ornithischia belong the bipedal ornithopods or bird footed dinosaurs the quadrupedal stegosaurs the heavily armored ankylosaurs the cera [illegible]

as the Late Jurassic *Camptosaurus* an animal reaching 4 m in length bipedal but with well developed forelimbs on which it may have stood while feeding The rear of the jaws bore a single row of blunt ridged serrate edged teeth *Iguanodon* from the Early Cretaceous of Europe and one of the first dinosaurs to be discovered stood 4 3 m tall

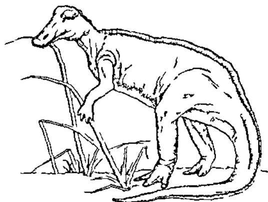
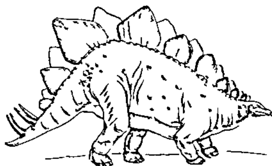
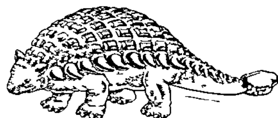


Fig 7 Restoration of the duckbilled dinosaur *Anatosaurus (Trachodon)* from the Late Cretaceous of North America these were 30–40 ft in length



Fg 8 Restorat on of the armored Jurass c d nosaur *Stegosaurus* (about 20 ft long) (*After C W G lmore*)



Fg 9 Restorat on of the armored Cretaceous d nosaur *Ankylosaurus* (20 ft long approx)

and was nearly 8 m long its jaws contained several rows of teeth similar to those of *Camptosaurus* A stout spike on the hand takes the place of the first toe or thumb its use is unknown

In the Late Cretaceous of North America and eastern Asia the duck billed d nosaurs were the most common herbivores *Trachodon* more correctly known as *Anatosaurus* was a large biped similar to *Camptosaurus* and *Iguanodon* but with the front of its mouth widened into a d icklike l eak suitable for scooping up water plants on which it fed (Fig 7) Mummified specimens of these dinosaurs reveal that the toes were webbed for swimming In the jaws were as many as 2000 teeth arranged in several rows to form an efficient grinding surface not unlike that of a horse s cheek teeth

Closely related to *Trachodon* were a number of other duck bills with extraordinary helmetlike or spikelike crests on their skulls formed l y extensions of the nasal passages These may possibly have functioned in some unknown fashion to aid the animal in feeding under water

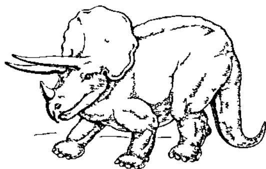
The stegosaurs or armored dinosaurs were ungainly quadrupedal animals with extremely long rear and short front legs a series of large triangular plates or spikes in alternate rows running down the back and long sharp spikes near the end of the tail The head was extremely small and the dentition feeble *Scelidosaurus* a primitive stegosaur from the Early Jurassic of England is the oldest well known ornithischian *Stegosaurus* (Fig 8) was from the Late Jurassic of North America related genera are known from the Jurassic of Europe and Africa

In the Cretaceous another type of armored dinosaur with shorter legs and wide flattened bodies replaced *Stegosaurus* These ankylosaurs have been called the horned toad dinosaurs because of a superficial resemblance to the lizard *Phrynosoma* *Hylaeosaurus* from the Lower Cretaceous of England was one of the animals originally termed dinosaur l y Owen *Polacanthus* also Early Cretaceous had l road paired spines over the anterior part of its l ody and a shieldlike buckler of small plates over the hips *Ankylosaurus* (Fig 9) and others from Late Cretaceous of North America and Mongolia had the entire upper surface of the body and head armored with thick bony plates that were drawn out into spikes along the sides

The horned dinosaurs or ceratopsians are confined to the Late Cretaceous of Mongolia and North America but appear to be descended from a deep-faced ornithopod such as *Psittacosaurus* from the Mongolian Early Cretaceous These quadrupeds had short necks and tails large heads with a deep parrotlike beak and a bony crest or frill extending far back over the neck *Protoceratops* from Mongolia the oldest ceratopsian was less than 2 m long and hornless The North American *Monoclonius* had a long single horn over the nose *Triceratops* largest (5 m long) and last of the group had long horns over the eyes and a shorter one on the nose (Fig 10) *Pachyrhinosaurus* bore a single immense horn in the middle of the forehead

Among the most curious dinosaurs is the small *Stegoceras* (Fig 11) whose skull about 20 cm long has a solid rounded mass of bone 10 cm thick above the minute brain cavity whence the family name Pachycephalosauridae Its dentition was feeble resembling that of armored dinosaurs little is known of the rest of the skeleton Related forms all from the Cretaceous of North America attained twice its size

Extinction of dinosaurs At the end of the Mesozoic era the dinosaurs and many other kinds of reptiles including the flying pterosaurs and the marine mosasaurs ichthyosaurs and plesiosaurs became extinct Certain invertebrate animals too notably the ammonites likewise became extinct at this time The reasons for the extinction of a relatively large number of diverse organisms in certain epochs of geologic time and the reasons for the



Fg 10 Restorat on of the horned d nosaur *Tr ceratops* (about 25 ft long) Late Cretaceous North Amer ca

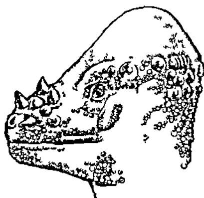


Fig 11 Head of *Stegoceras* from Late Cretaceous of Canada The rounded dome above the level of the eyes is solid bone

disappearance of orders containing a variety of highly successful adaptive types are among the most perplexing problems of paleontology

Inability to adapt to a changing environment is the usual cause of extinction Adaptation is a complex interrelationship between the structures physiological processes and behavior of an organism and many different aspects of its environment including other organisms Changes in the environment induce changes in a population of animals but if these changes are not in a direction which is adaptive to the new conditions or if they do not occur rapidly enough to keep the animals adapted to the shifting environment extinction will occur

Under such conditions of change those animals with the narrowest specializations are most apt to be wiped out A broad adaptation to several types of food and habitat may enable one kind of ani

[illegible]

Late Cretaceous extinctions may well be the result of their habits

When the changing physical conditions drastically reduce the numbers of one kind of animal many others may be affected because the food chains on which all animals *depend for* nourishment and of which each in turn forms a part may be disrupted Thus one extinction may set off a whole chain of exterminations

It has been suggested that upheaval of mountains in western North America at the end of the Mesozoic era drained the swampy lowlands on which the larger dinosaurs lived and thus led to their extinction Local upheavals might have wiped out suitable habitats in some regions but these changes were so slow that the animals involved could have migrated to other more favorable places The near simultaneous disappearance of marine animals as well cannot be explained in this way Disease has often been suggested but cannot be demonstrated Extremes of temperature are not supported by the fossil record of Late Cretaceous and Paleocene plants which indicate a widespread warm temperate climate Greatly increased radiation of some type (such as cosmic rays) also not demonstrable which might have rendered the dinosaurs and other forms sterile should have affected all kinds of organisms and not merely selected certain orders for its lethal activity In short it is not possible to point to the particula environmental factor which led to dinosaurian (o other) extinction It is known that the end of th Mesozoic was a time of many changes in land an sea some combination of which rendered the worl unsuitable for these great beasts which had domi nated it for longer than 100 000 000 years [J T C

Bibliography E H Colbert *The Dinosaur Book* 1951 O C Marsh *The Dinosaurs of North America* USGS 16th Ann Rept, 1896

Dioctophymoidea

A group of nematodes characterized by the peculiar structure of the copulatory bursa of the male Various species are parasites of fish aquatic birds and mammals This group is considered to constitute either an order or superfamily according to the viewpoint of the specialist

Dioctophyme renale This species the so-called giant kidney worm is the best known in the group It is a fairly common parasite of dogs and a wide variety of mammals especially the mink Knowledge of the life cycle depends almost entirely on the studies reported by A Woodhead The adult worms are among the longest nematodes the females measuring up to 1 meter The size blood red color and characteristic location in the kidney led to the common name Sometimes found in the abdominal cavity they probably penetrate the kidney before they become mature They live in the kidney for 1 3 years digesting its substance until it is only a shell The infection may be fatal but is often symptomless The eggs pass in the urine hatch only when swallowed 6–12 months later by branchiobdellid annelid worms which live attached to the outside of crayfish After a period of development they encyst in the annelid and if this host is eaten by certain fish they resume development and again encyst Development is completed in mammals which eat the fish the entire cycle requiring about 2 years *See* OLIGOCHAETA

Other genera Other genera include *Eustrongylides* and *Hyrstrichis* which inhabit the proventriculus of aquatic birds and are apparently transmitted by fish Larval forms have been the object of studies in physiology by T Von Brand Species of *Soboliphyme* are parasites in the intestine of foxes cats and the wolverine *See* NEMATODA [J A S]

Bibliography T Von Brand *Biol Bull* 92 162 166 1947, A E Woodhead *Trans Am Microscop Soc* 69 21–46 1950

Diode, semiconductor

A two-terminal device that utilizes the rectifying properties of a semiconductor material to convert an alternating current into a pulsating direct current Germanium and silicon are the semiconduc

tor materials most commonly used in diodes In a junction diode the rectification is produced at the junction between p type and n type material by a phenomenon similar to that occurring in transistors In a point contact diode, rectification occurs at the contact between a sharp point and a pellet of semiconductor material The commonest types of semiconductor diode have axial wire leads emerging through the sealed ends of a tiny glass cylinder containing the rectifying element These diodes are widely used in computers and in television and radio receivers *See* JUNCTION DIODE, POINT CONTACT DIODE, SEMICONDUCTOR [J MR]

Diode, vacuum

The diode vacuum tube is the simplest of all tubes It contains a cathode and an anode in an evacuated envelope The cathode in addition has within it a filament to heat its surface and facilitate electron emission A potential which is positive relative to that of the cathode is connected to the anode so that current in the form of an electron stream is drawn from the cathode to the anode

Physical construction Vacuum diode envelopes are commonly made of glass with a miniature or octal base which can be plugged into a socket Tubes range in size from ¼ in to over 1 in in diameter They usually have either an oxide coated cathode or an oxide coated filament which emits electrons directly Anodes or plates are commonly made of nickel which is blackened to increase the heat radiation

Cathode The negative terminal of an electron tube is called the cathode Such a terminal or electrode invariably is an electron emitter Electrons possessing a negative charge are accelerated from a negative to a positive electrode by the electric field between the electrodes

Cathodes are of several types the commonest is the thermionic cathode in which the emission of electrons is produced by heating the cathode This can be done either by passing current directly through it or by heating it indirectly from a filament Thermionic cathodes are commonly made of material of a low work function This means that electrons require relatively little energy to overcome the surface potential barrier which exists at the surface of the cathode in passing from the conducting region within the cathode to the vacuum or low pressure gas outside *See* ELECTRON EMISSION

Cathodes may also be of the cold emitting variety Such cathodes are also coated with a low work function material but generally operate in a low pressure gas rather than in a vacuum Current initiation is aided by a probe or point on the cathode about which a high gradient of potential develops as the potential between electrodes is raised

Thermionic cathodes are capable of producing current densities up to several amperes per square centimeter Cold cathodes are ordinarily not able to produce more than several milliamperes per square centimeter Thermionic cathodes in gas tubes give rise to the highest of all densities which may be of the order of tens of amperes per square centimeter Because of the dependence of electron tubes upon electron and ion flow, every electron tube must have a cathode *See* ELECTRICAL CONDUCTION IN GASES ELECTRON MOTION IN VACUUM

Filament Conductors used to heat indirect cathodes of thermionic tubes are called filaments They are also used directly as electron emitters in some types of vacuum tube Such filaments are invariably made of tungsten since it has the highest melting temperature of all of the metals (3370°C) and also a low vapor pressure (it does not evaporate readily) Where filaments are used for indirectly heated cathodes they must be covered with an insulating coating which can stand the high operating temperature of the filament This insulation is usually an aluminum oxide

Where filaments are used as emitters or cathodes they are commonly covered with low work function materials such as barium and strontium oxides These materials emit electrons readily at relatively low temperatures corresponding to a dull red heat

Filaments undergo considerable expansion when they are heated to produce emission Therefore careful attention must be paid to design to avoid mechanical strains or failures from expansion and contraction In addition the electrical resistance of filaments undergoes considerable change when they are heated to produce emission The cold resistance of a filament is quite low but it can attain a value 5–10 times its cold resistance when heated to

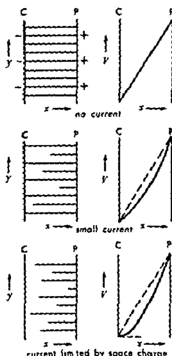


Fig 1 Electric flux lines and potential distribution in a plane-electrode diode for various degrees of space charge (From K R Spangenberg Fundamental Electron Devices, McGraw Hill 1957)

electron emission temperatures For this reason attention must sometimes be paid to circuits which control the current rather than the voltage across the filaments

Anode or plate The plate of an electron tube is also known as the anode The plate is always operated positive relative to the cathode The term plate is used because the shape of the anode is commonly that of a box of flat plates but it may also be a circular cylinder

Plates or anodes of electron tubes must be made of a conducting material which can be raised to a high temperature in a vacuum or a gas without introducing any harmful effects or being physically damaged Plates are commonly made of some metal with a high melting temperature and a low vapor pressure such as molybdenum tantalum or nickel (in receiving tubes) occasionally they are made of carbon or graphite The plates must not evolve gas when they are bombarded by electrons sometimes plates are covered with a material that will absorb any gas in the tube Such materials include zirconium and tantalum

In small electron tubes the plates are radiation cooled Such plates will be raised to a red heat when they are dissipating several watts per square centimeter Above this power density the plate structure is weakened thus a limit is set to the power the tube can handle Where higher power densities are required the plates are water cooled In contrast to radiation cooled plates which can dissipate 4–10 watts/cm² water cooled plates can dissipate 30–110 watts cm² Water cooled plates are commonly the vacuum envelope of the tube and water is circulated over their outer surface Such plates can also be covered with radiating fins for air cooling

Theory of operation The theory of the vacuum diode is relatively simple The tubes usually have either a plane or a cylindrical geometry In the plane-geometry case the cathode and the anode are essentially parallel plates between which when the cathode is cold the potential will rise linearly from the cathode to the plate as it does in a simple parallel plate capacitor However when the cathode is heated so that it emits electrons, these electrons in motion from cathode to anode produce a negative space charge which depresses the potential between the electrodes This can occur only until the potential gradient at the cathode has been reduced to zero as shown in Fig 1 The result of this action is that the current density J between the electrodes is proportional to the $\frac{3}{2}$ power of the potential difference V and inversely proportional to the distance x squared

$$J = AV^{3/2}/x^2$$

This relation is known as the Langmuir Child law

However the law expressed above is an ideal one and it breaks down at high and low voltages as shown in Fig 2 At low voltages the current is influenced by the velocity of emission of the electron so that current may flow even when the voltage is negative This is the case in the region marked as A The $\frac{3}{2}$-power law of voltage variation will hold over a considerable range of the characteristics as shown by the region marked B However at large voltages emission saturation occurs so that the current tends to be limited relative to further increases in voltage This is the case in the region marked C Two kinds of saturation may be distinguished in diodes There is first the voltage saturation already referred to and shown in Fig 3 for various cathode temperatures If the current is displayed as a function of cathode temperature for various voltages there is exhibited what is known as temperature saturation as shown in Fig 4

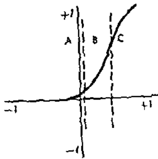


Fig 2 Voltage-current characteristics of a vacuum diode (From K R Spangenberg Fundamentals of Electron Devices McGraw Hill 1957)

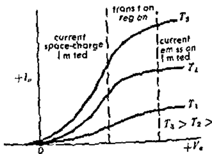


Fig 3 Current voltage characteristics of a vacuum diode for various cathode temperatures showing voltage saturation (From K R Spangenberg Fundamentals of Electron Devices, McGraw Hill 1957)

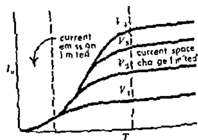


Fig 4 Current–cathode-temperature characteristics of a vacuum diode for various anode-to-cathode voltages showing temperature saturation (From K R Spangenberg Fundamentals of Electron Devices McGraw Hill 1957)

The space-charge limitation of current referred to above is a significant one and determines the amount of current that can be drawn in a diode To overcome this limitation the cathode anode spacing is generally made as small as possible The current is then still limited for a given voltage by the Langmuir Child law up to the point where the current called for by the law is greater than the cathode can supply

Applications of diodes One of the commonest uses of diodes is for detection of radio frequency (rf) signals *See* DETECTOR

Diodes are also used extensively as rectifiers in power supply circuits In these circuits the diode acts as a one way (unilateral) element which permits the current to flow in one direction but not in the reverse direction *See* RECTIFIER

Because of their nonlinear characteristics diodes may also be used as modulators that is to put an audio signal on an rf carrier If an audio and an rf signal are applied to a diode at the same time the output of the diode consists of the rf signal whose amplitude follows (is modulated by) the audio signal *See* MODULATION MODULATOR

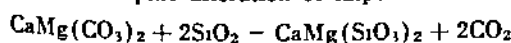
Diodes are also used extensively in computers and for various wave-shaping applications They may be used to limit the amount of current or voltage that can be developed or to sharpen pulses Probably the most common use is as switching elements which pass current when the anode is positive relative to the cathode but not when the anode is negative relative to the cathode Most electronic computer circuits will contain two or three times as many diodes as they do amplifying elements Sometimes two complete diode structures are put within the same vacuum envelope Such tubes are known as duodiodes They find many uses as full wave rectifiers detectors and computer switches [K R S]

Diopside

The name given to the monoclinic pyroxene $CaMg(SiO_3)_2$ Diopside is one end member of a complete solid solution series with the iron end member hedenbergite and forms a limited solid solution series with the enstatite ferrosilite series except at temperatures near the melting point where solid solution is complete At 700°C less than 4% of the enstatite end member is in solid solution with diopside Diopside is gray to white short stubby prismatic often equidimensional crystals with the 87° pyroxene (110) cleavages Small amounts of iron impart a greenish color to the minerals The indices of refraction increase with increasing iron for pure diopside they are $n\alpha = 1\,664$ $n\beta = 1\,671$ $n\gamma = 1\,694$ Pure diopside is common and occurs as a metamorphic alteration of impure dolomites

$$CaMg(CO_3)_2 + 2SiO_2 - CaMg(SiO_3)_2 + 2CO_2$$

in medium and high grades of metamorphism Diopside is associated with such minerals as calcite quartz tremolite garnet plagioclase olivine and orthopyroxene Diopside also occurs in many ba



salts perdiotites dunites and other magnesium rich rocks and in slags and some meteorites *See* AUGITE, ENSTATITE PYROXENE [C W D]

Diopter

A measure of the power of a lens or a prism The diopter (also called dioptrie) is usually abbreviated D Its dimension is a reciprocal length and its unit is the reciprocal of 1 meter Thus a thin lens of κ diopters has a focal length of $1000/\kappa$ mm or $39\,4/\kappa$ in The lens is collecting for positive κ diverging for negative κ and afocal for $\kappa = 0$ *See* FOCAL LENGTH LENS OPTICAL

One can speak of the power of a single surface The power of a lens is then the sum of the powers of its surfaces in diopters Analogously the power of a group of (thin) lenses in contact is the sum of the powers of the single lenses

For a lens which is doubly symmetric (having toric or cylindrical surfaces for instance) two powers—one maximal and one minimal—must be assigned These correspond to the powers in two perpendicular planes

The power of a prism is the measure of the deviation of a ray going through a prism measured at the distance of 1 meter A prism that deviates a ray by 1 cm in a distance of 1 meter is said to have a power of one prism diopter *See* PRISM OPTICAL

Spectacle lenses in general consist of thin lenses which are either spherical to correct the focus of the eye for near and far distances or cylindrical or toric to correct the astigmatism of the eye An added prism corrects a deviation of the visual axis The diopter thus gives a simple method for prescribing the necessary spectacles for the human eye *See* EYE GLASSES [M H]

Diorite

A phaneritic (visibly crystalline) plutonic rock with granular texture composed largely of plagioclase feldspar (oligoclase or andesine) with smaller amounts of dark colored (mafic) minerals (hornblende biotite or pyroxene) This dark gray rock is used occasionally as ornamental and building stone and is known commercially as black granite For a general discussion of textural structural and compositional characteristics *see* IGNEOUS ROCKS

Mineralogy Gray or white feldspar grains commonly show on broken surfaces fine parallel striations or twin lines Under the microscope plagioclase feldspar exhibits striking examples of zonal structure in which individual grains are composed of concentric shells of rectangular outline and differing composition Most commonly internal shells are calcium rich more external shells are more sodic Not uncommonly the change in composition of successive shells from center to exterior may be reversed interrupted oscillatory or repetitive As the plagioclase feldspar becomes more calcic than andesine the rock passes into gabbro With increase in plagioclase content

decrease in other constituents diorite passes into anorthosite

One or more mafic minerals may be present Hornblende (microscopically green or brown) as irregular or elongate grains is the most common Black flaky biotite mica (microscopically brown) may be intergrown with hornblende It is most abundant in quartz rich diorites Augite is uncommon and may be converted to hornblende Olivine and orthopyroxene are rare

Quartz, generally interstitial to plagioclase may be present in small amounts Where it constitutes over 5% of the minerals the rock is called quartz diorite (tonalite) Small amounts of potash feldspar closely associated or intergrown with quartz (micropegmatitic texture) are common but where present in excess of 5% the rock is a monzonite Nepheline and other feldspathoids are rare constituents Magnetite ilmenite and apatite are the most common accessory minerals but zircon and sphene occur in certain varieties

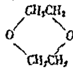
Textures and structures Although generally equigranular the texture may be porphyritic with large crystals (phenocrysts) of plagioclase or hornblende Porphyritic types may grade into diorite porphyry Some diorites and quartz diorites (tonalites) carry orbicular structures and many show flow structures and banding

Occurrence Diorite occurs as isolated small bodies such as dikes sills and stocks In more irregular forms it may be associated with granodiorite and granite or with gabbro It occurs most abundantly in orogenic (fold mountain) belts

Origin Diorite forms in many different ways Some has crystallized directly from a dioritic magma (rock melt) Some is of hybrid origin and formed by reaction between a magma and contaminating foreign rock fragments (*xenoliths*) Many diorites are products of solid state transformation or metasomatism Gabbro may be converted to diorite by a relative loss in calcium iron and magnesium and a gain in sodium and silicon See MAGMA METASOMATISM PETROGRAPHIC PROVINCE. [C A CA]

Dioxane

The cyclic ether of ethylene glycol It is also called 1 4-dioxane and p-dioxane Originally a by product of ethylene glycol manufacture it is now also prepared by the catalytic dimerization of ethylene

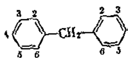


oxide It is unusual among substances of low dielectric constant (2 21) in that it is soluble in water in all proportions Extensively used as a solvent industrially it readily dissolves fats waxes natural and synthetic resins cellulose ethers and lacquers and it is employed by biologists to prepare paraffin impregnated tissue sections It reacts readily with chlorine to give dichlorodioxane a reactive intermediate for the manufacture of plasticizers and insecticides The hydrosulfate which melts at 101°C is typical of the unusual stability of many additives. Because dioxane causes some liver damage, excessive breathing of the vapor or exposure of the skin to the liquid should be avoided A peroxide which forms easily in crude dioxane detonates on distillation The boiling point of dioxane is 101 3°C melting point 12 5°C refractive index 1 4224 specific gravity 1 03.6 [B X S]

See ETHER ETHYLENE OXIDE

Diphenylmethane

A colorless hydrocarbon $(C_6H_5)_2CH_2$ melting point 25 9°C boiling point 263 2°C, described as having an odor resembling that of geraniums or oranges It can be prepared by the action of formaldehyde (HCHO) on benzene in the presence of



Diphenylmethane

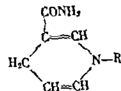
sulfuric acid by the action of benzyl chloride $C_6H_5CH_2Cl$ on benzene in the presence of aluminum chloride or by the reduction of benzophenone $C_6H_5COC_6H_5$

Diphenylmethane may be hydrogenated in the ... lmethane

Although diphenylmethane has been us... perfume for soaps it is not an important item of commerce A derivative 2 2 dihydroxy 3,5 6 3,5 6 hexachlorodiphenylmethane has been used as a bacteriostatic agent in soaps See POLYNUCLEAR HYDROCARBON [C.R R]

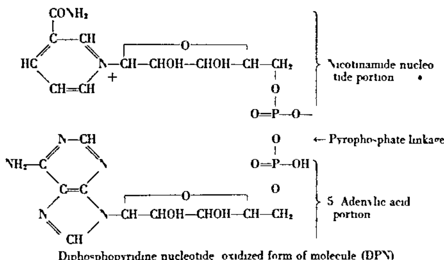
Diphosphopyridine nucleotide (DPN)

An organic coenzyme and one of the most important components of the enzymatic systems concerned with biological oxidation reduction reactions It is also known as DPN diphosphopyridine dinucleotide coenzyme I and codehydrogenase I (see COENZYME) DPN is found in the tissues of all living organisms



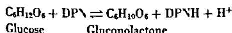
Diphosphopyridine nucleotide reduced form of nicotinamide portion in DPNH

The nicotinamide or pyridine portion of DPN can be reduced chemically or enzymatically with the formation of reduced or hydrogenated DPN (DPNH) DPN functions as the immediate oxidizing agent for the oxidation or dehydrogenation of



Diphosphopyridine nucleotide oxidized form of molecule (DPN)

various organic compounds in the presence of appropriate dehydrogenases which are specific apoenzymes or protein portions of the enzyme In the dehydrogenase reactions one hydrogen atom is transferred from the substrate to DPN while an other is liberated as hydrogen ion For instance in a reversible reaction catalyzed by glucose dehydrogenase, glucose is oxidized to gluconolactone as follows



$$C_6H_{12}O_6 + DPN \rightleftharpoons C_6H_{10}O_6 + DPNH + H^+$$

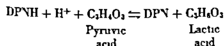
Glucose Gluconolactone

The DPNH formed in biological oxidations is reoxidized to DPN in coupled reductions also catalyzed by specific enzymes In respiration the DPNH is reoxidized through a sequence of reactions in which a flavoprotein enzyme diaphorase and the cytochrome (*see* CYTOCHROME) system of iron porphyrin catalysts transfer the electrons from DPNH to molecular oxygen the over all reaction being



$$2DPNH + 2H^+ + O \rightarrow 2DPN + 2H\ O$$

In fermentations DPNH is reoxidized with the concomitant reduction of organic molecules which are usually produced in intermediary metabolism For example in the metabolism of muscle tissue lactic acid is produced by the reduction of pyruvic acid



$$DPNH + H^+ + C_3H_4O_3 \rightleftharpoons DPN + C_3H_6O_3$$

Pyruvic acid Lactic acid

The enzyme catalyzing this reaction is called lactic dehydrogenase since the process is reversible and lactic acid can be oxidized with DPN DPN and its reduced form DPNH serve to couple oxidative and reductive processes and are constantly regenerated during metabolism Hence they serve as catalysts and DPN is referred to as a coenzyme In some enzymatic reactions a different coenzyme triphosphopyridine nucleotide (*see* TRIPHOSPHOPYRIDINE NUCLEOTIDE (TPN)) or coenzyme II is required Dehydrogenases are generally quite specific with respect to the coenzyme which they can use *See* BIOLOGICAL OXIDATION ENZYME NIACIN

[M D]

Diphtheria

A communicable disease of man caused by the growth of the bacterium *Corynebacterium diphtheriae* on the mucous membranes or less commonly in cutaneous tissues.

Pathogenesis Both toxigenic and nontoxigenic strains of *C diphtheriae* are capable of inducing the formation of a spreading pseudomembranous exudate which may cause serious obstruction in the larynx and trachea The classic pattern of the disease caused only by the toxigenic strains, involves the spread of diphtheria toxin to target tissues far from the superficial site of the infection The toxin is a protein highly toxic for most animal with a molecular weight of 70 000 While the di ease i often fatal cases of varying degrees of severity occur Recovery is followed by immunity produced by circulating antibody

Immunization and Schick test Although diphtheria has been of world wide occurrence in recent years its incidence has declined sharply in countries where there has been a general immunization of children with diphtheria toxoid a toxin rendered nontoxic by mild treatment with formaldehyde The immune status of an individual with respect to diphtheria toxin may be determined through the use of the modified Schick test Approximately 0 0025 micrograms of toxin protein or one-fiftieth of the minimal amount required to kill a guinea pig weighing 250 g is injected into the skin of the forearm the test site An equal amount of toxoid is injected at a control site Necrosis at the test site indicates a nonimmune state while immunity the presence of circulating antibody is indicated by a lack of reaction at either site Immunity complicated by allergy to diphtherial proteins results in a delayed inflammatory reaction at both sites

C diphtheriae *C diphtheriae* in gram stained preparations from inspissated serum slants in

Loeffler's medium in blood agar or chocolate tellurite agar offers a multiplicity of shapes. They are irregular gram-variable rodlike forms which are usually tapered but often club shaped. In old cells certain basic dyes such as alkaline methylene blue and toluidine blue become localized on intracellular spherical polyphosphate-containing inclusions rendering them red. These are the so-called metachromatic granules. Actively growing diphtheria bacilli do not contain detectable metachromatic granules but appear as a population of uniform gram-positive rods tapered at one end. The actual size of the bacilli varies from one strain to another. Three colonial types reflecting differences in cell size, shape, and surface properties are recognized: the smooth, dwarf smooth, and rough. *See* CORYNEBACTERIACEAE.

Methods for the specific identification of diphtheria bacilli by serologic and phage typing, while known, are not yet in general use. Attempts have been made to establish a correlation between the colonial types isolated from patients and the clinical severity of diphtheria. Thus smooth strains associated with mild infections are designated *mitis*, rough strains from severe infections are designated *gravis*, and dwarf smooth strains are classed as *intermedius*. *C. diphtheriae* has been further differentiated on the basis of starch fermentation, a property correlated with a high degree of invasiveness. As already indicated, invasiveness and toxigenicity are separate properties which do not necessarily occur together in a single strain. Because of this fact the use of *gravis*, *mitis*, and *intermedius* is not always a meaningful classification. For example, an invasive *mitis* strain which is toxigenic would produce infection of more serious consequence than an invasive nontoxigenic *gravis* strain. While little is understood about the genetic basis for invasiveness in *C. diphtheriae*, considerable is known about the genetic control of toxigenicity. Nontoxigenic strains which are rendered lysogenic following infection with certain temperate bacteriophages become toxigenic. Toxin production appears to be directly under the genetic control of the latent or carried bacteriophage. Elimination of the viral genetic element or provirus from the cell results in a return to the nontoxigenic state. *See* BACTERIOLOGY, MEDICAL; BACTERIOPHAGE; SEROLOGY. [L.D.]

Diphyllidea

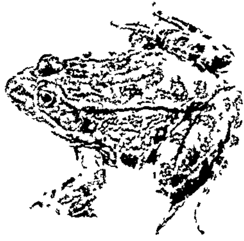
An order of tapeworms of the subclass Cestoda. Species of this order belong to a single genus and live in the intestine of elasmobranch fishes. The scolex has a large muscular rostellum armed with hooks and two fused pairs of suckers; the neck is armed with T-shaped hooks. The segments are peculiar in having the genital atrium in the midline and in the arrangement of the genital organs. Larval stages have been found parasitizing marine mollusks and crustaceans, but the history is not known. *See* CESTODA; *see also* TETRAPHYLLIDEA. [C.P.R.]

Diplasiocoela

A suborder of frogs of the order Salientia in which the eighth vertebra is biconcave. The remaining presacral vertebrae are procoelous and the sacral vertebrae articulate with the coccyx by a double condyle. The pectoral girdle is firmisternal; the right and left halves of the girdle are firmly joined at the midline, not overlapping as in most other frogs. This is the largest order of frogs, with about 1200 species divided among three families: Ranidae, Rhacophoridae, and Microhylidae.

Ranidae. The Ranidae number about 470 species, of which approximately 170 belong to the genus *Rana* (see illustration). This genus is fairly well represented wherever frogs live except in South

to be filled by leptodactylids (see FROG). The genus *Rana* includes the abundant typical frogs as opposed to toads of North America and Europe. In Africa and Asia, in addition to *Rana*, there are more than 30 other genera of diverse habits and appearance that belong to the Ranidae.



The leopard frog (*Rana pipiens*), the most widespread species in North America, found from Hudson's Bay to Panama and from the Atlantic almost to the Pacific. (*American Museum of Natural History*)

Rhacophoridae. In the suborder Diplasiocoela the family that corresponds to the tree frog family Hylidae of the Procoela is the Rhacophoridae. Externally hylids and rhacophorids may resemble one another very closely, but the characteristics of the vertebral column and pectoral girdle indicate their true relationships. The geographic distribution of the two families is largely mutually exclusive. Where hylids are abundant, in the New World, Australia, and New Guinea, there are no rhacophorids. Over 400 species of Rhacophoridae are credited to the fauna of Africa, where only a single *Hyla* is known. Similarly, about five species of *Hyla* occur in Asia among over 100 rhacophorids. The

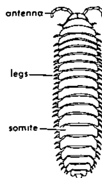
center of rhacophorid diversity is Africa, where about 20 genera are found Only two genera and these are doubtfully distinct, occur in Asia

Microhylidae. The frogs of the family Microhylidae are found in three separate regions of the world southern Africa and Madagascar Ceylon India and southern China to New Guinea and the extreme northern part of Australia and South America northward into the southern United States Many of the microhylids are chunky little frogs with a pointed head and tiny mouth that lead a terrestrial or burrowing life In some regions most notably in New Guinea, extensive adaptive radiation has produced a variety of ecological types paralleling more typical members of other families The specialization of microhylids is such that there are more genera recognized (58) than of ranids and rhacophorids together, though in number of species (about 230) the microhylids are inferior to either of the other groups *See* AMPHIBIA SALIENTIA [R C Z]

Diplopoda

A class of terrestrial, tracheate, oviparous arthropods Diplopods (millipeds) are largely cryptic in habits are saprophytic feeders and are characterized by the development of a compact head with a pair of short, simple, 8 jointed antennae powerful mandibles and a subbuccal gnathochilarium formed from embryonic maxillary elements The body is not differentiated into thorax and abdomen but is composed of a variable number of similar cylindrical diplosomites, each of which (except the first two or three) bears two pairs of walking legs The body wall is chitinous, with a thick impregnation of calcium carbonate in the majority of species Most millipeds are variously adapted for rolling into a closed spiral or nearly perfect sphere when threatened Typically each abdominal segment is provided with a pair of glands which effuse a volatile poisonous fluid Respiration occurs through profuse fine tracheae opening through stigmata near the bases of the legs The digestive tract is a straight tube, except in oniscomorph species in which it is somewhat coiled The anal opening is in the last segment and is closed by two tightly fitting anal valves Ducts from the reproductive organs pass through or behind the second pair of legs to the external openings

Reproduction The sexes are separate and fertilization is internal, following prolonged clasping behavior Spermatic masses are extruded beforehand from the male seminal opening onto the gonopods (modified legs of the seventh segment) from which sperm material is transferred into seminal receptacles in the cyphopods (specialized structures terminating the outer ends of the oviducts) In oniscomorph millipeds the gonopods are not developed and the male transfers spermatophores with his mouthparts Eggs vary greatly in size and number They may be laid in a cluster and "brooded" by the mother, scattered singly in the humus environment, or enclosed in an igloo-shaped mud nest built by the mother Postembryonic development is anamorphic, and the individual passes through seven growth stages with segments and legs added at each molting period Development is gradual, without major changes in appearance, and may require a year or more for completion Mating usually takes place soon after the molt into sexual maturity



Diplopoda (*From R E Snodgrass, A Textbook of Arthropod Anatomy, Cornell Univ Press, 1952*)

At the present more than 8000 species have been described although so far only the fauna of Europe is well known Existing classifications are still unsatisfactory but in general about 11 orders and more than 111 families are recognized Probably as many as 25 000 species will eventually be recognized Classification is based to a large extent upon shape of the male gonopods which are quite constant and characteristic for each species For higher groups the forms of segments mouthparts and legs are useful

Most species are local in distribution because millipeds generally remain close to the parental habitat and some may be restricted to a few square miles Even genera are limited in distribution, and only a few occur on more than one continent This endemism plus dependence on a continuously moist and undisturbed habitat makes diplopods well suited for studies of zoogeography and evolution

Classification and distribution. The diplopods have a long geological history they arose in the early Devonian and were well developed by the late Pennsylvanian There has been little adaptative radiation in the Diplopoda, and less external modification of body form than in perhaps any other large class of arthropods *See* ARTHROPODA [R L HO]

Diploporita

An order of Cystidea in which the thecal canals were associated in pairs, the canals ran perpendicularly to the surface and, when exposed in fossils by superficial abrasion, appear as paired pores termed diplopores Exposure of solitary canals produces haplopores These so-called pores are, of course artifacts The theca was ovoid or pear shaped, built up of numerous small polygonal plates arranged without order There was usually no stem Early forms such as *Aristocystites* in the

Loeffler s medium in blood agar or chocolate tellu rite agar offers a multiplicity of shapes They are irregular gram variable rodlike forms which are ... h shaped In old cells

lar spherical polyphosphate containing inclusions rendering them red These are the so called meta chromatic granules Actively growing diphtheria bacilli do not contain detectable metachromatic granules but appear as a population of uniform gram positive rods tapered at one end The actual size of the bacilli varies from one strain to an other Three colonial types reflecting differences in cell size shape and surface properties are recognized the smooth dwarf smooth and rough *See* CORYNEBACTERIACEAE

Methods for the specific identification of diph theria bacilli by serologic and phage typing while known are not yet in general use Attempts have been made to establish a correlation between the colonial types isolated from patients and the clini cal severity of diphtheria Thus smooth strains associated with mild infections are designated *mitis* rough strains from severe infections are des ignated *gravis* and dwarf smooth strains are classed as *intermedius* *C diphtheriae* has been further differentiated on the basis of starch fermen tation a property correlated with a high degree of invasiveness As already indicated invasiveness and toxigenicity are separate properties which do not necessarily occur together in a single strain Be cause of this fact the use of *gravis mitis* and *in termedius* is not always a meaningful classification For example an invasive *mitis* strain which is toxigenic would produce infection of more serious consequence than an invasive nontoxigenic *gravis* strain While little is understood about the genetic basis for invasiveness in *C diphtheriae* consider able is known about the genetic control of toxi genicity Nontoxigenic strains which are rendered lysogenic following infection with certain temper ate bacteriophages become toxigenic Toxin pro duction appears to be directly under the genetic control of the latent or carried bacteriophage

PHAGE SEROLOGY [L B]

Diphyllidea

An order of tapeworms of the subclass Cestoda Species of this order belong to a single genus and live in the intestine of elasmobranch fishes The scolex has a large muscular rostellum armed with hooks and two fused pairs of suckers the neck is armed with T shaped hooks The segments are pe culiar in having the genital atrium in the midline and in the arrangement of the genital organs Lar val stages have been found parasitizing marine mollusks and crustaceans but the history is not known *See* CESTODA *see also* TETRAPHYLLIDEA [C P R]

Diplasiocoela

A suborder of frogs of the order Salientia in wh h the eighth vertebra is biconcave The rema n ng presacral vertebrae are procoelous and the sacral vertebrae articulate with the coccyx by a double condyle The pectoral girdle is firmisternal the right and left halves of the girdle are firmly joined at the midline not overlapping as in mo t o her frogs This is the largest order of frogs with about 1200 species divided among three families Ran dae Rhacophoridae and Microhylidae

Ranidae The Ranidae number about 470 spec es, of which approximately 170 belong to the genus *Rana* (see illustration) This genus is fairly well represented wherever frogs live except in South ... cont nents

to be filled by leptodactylids (see PROCOEL ... ope ... are ... and more than 30 other genera of diverse ... appearance that belong to the Ranidae



The leopard frog (*Rana p p ens*) the most widesp ead spec es in North Amer ca found from Hudson s Bay to Panama and from the Atlant c almost to the Pacific (*Amer can Museum of Natural H story*)

Rhacophoridae In the suborder Diplasiocoela the family that corresponds to the tree frog fam ly Hylidae of the Procoela is the Rhacophoridae Ex ternally hylids and rhacophorids may resemble one another very closely but the characteristics of the vertebral column and pectoral girdle indicate the r true relationships The geographic distribution of the two families is largely mutually exclus e Where hylids are abundant in the New World Australia and New Guinea there are no rhacopho rids Over 400 species of Rhacophoridae are cred ited to the fauna of Africa where only a s ngle *Hyla* is known Similarly about five species of *Hyla* occur in Asia among over 100 rhacophorids The

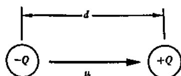
adapted to aerial respiration it does not estivate The Lepidosirenidae are represented today by *Lepidosiren paradoxa* in South America and four species of *Protopterus* in Africa These forms withstand desiccation by encasing themselves in a protective capsule of mud as swamps dry and they breathe atmospheric air until freed by seasonal rains *See* SARCOPTERYGII, *see also* CROSSOPTERYGII [R M B]

Bibliography A S Romer *Vertebrate Paleontology* 2d ed 1945

Dipole

Any object or system that is oppositely charged at two points or poles such as a magnet or a polar molecule The properties of a dipole are determined by its dipole moment that is the product of one of the charges by their separation directed along an axis through the centers of charge *See* DIPOLE MOMENT

An electric dipole consists of two electric charges of equal magnitude but opposite polarity separated by a short distance (see figure) or more



Electric dipole with moment $\mu = Qd$

generally a localized distribution of positive and negative electricity without net charge whose mean positions of positive and negative charge do not coincide For a discussion of electric dipole radiation *see* ELECTROMAGNETIC RADIATION

Molecular dipoles which exist in the absence of an applied field are called permanent dipoles while those produced by the action of a field are called induced dipoles *See* MOLECULAR ASSOCIATION POLAR MOLECULE [R D W]

The term magnetic dipole originally referred to the fact that a magnet has two poles and because of these two poles experiences a torque in a magnetic field if its axis is not along the magnetic flux lines of the field (*see* MAGNET) It is now generalized to include electric circuits which because of the current also experience torques in magnetic fields [R P WI]

Dipole moment

A mathematical quantity characteristic of a dipole unit equal to the product of one of its charges times the vector distance separating the charges (*see* DIPOLE) The dipole moment $\mathbf{\mu}$ associated with a distribution of electric charges q is given by

$$\mathbf{\mu} = \sum_i q_i \mathbf{r}_i$$

where $\mathbf{r}$ is the vector to the charge q For systems with a net charge (for example positive) the origin is taken at the mean position of the positive charges (and vice versa) Dipole moments have the dimensions coulomb meters in the rationalized mks system and statcouloumb centimeters in the cgs electrostatic system Molecular dipole moments are often expressed in debye units where 1 debye = 10^{-18} statcoulomb cm

The electric potential Φ of a dipole at a long distance R from the dipole is given by

$$\Phi = \gamma \frac{|\mathbf{\mu}| \cos \theta}{4\pi\epsilon_0 |\mathbf{R}|^2}$$

where θ is the angle between $\mathbf{\mu}$ and $\mathbf{R}$ ϵ_0 is the permittivity of vacuum and γ is a geometrical factor ($\epsilon_0 = 1$ and $\gamma = 4\pi$ in cgs electrostatic units, $\epsilon_0 = 8\,854 \times 10^{-12}$ farad/m and $\gamma = 1$ in rationalized mks units)

The potential energy U of a dipole in a uniform electric field E is $U = -|\mathbf{\mu}||\mathbf{E}| \cos \phi$ where ϕ is the angle between $\mathbf{\mu}$ and $\mathbf{E}$

The induced dipole moment $\mathbf{\mu}$ of a molecule may be expressed in terms of molecular parameters by the equation $\mathbf{\mu} = \alpha \mathbf{E}_L$ where α is the polarizability and $\mathbf{E}_L$ is the local field strength acting at the molecular site This relation permits the interpretation of the macroscopic polarization and hence dipole moments in terms of molecular processes *See* DIELECTRIC CONSTANT, POLARIZATION (DIELECTRICS) [R D W]

Diptera

An order of the class Insecta known as the true flies and so named because they possess only two wings This characteristic together with a pair of balancers or halteres distinguishes flies from all other orders of the Insecta Members of the order are known commonly as flies gnats or midges and these three names form a part of the common names of most families genera and species of true flies The term fly also forms a part of the compound names of the insects in many other orders as butterfly May fly and chalcid fly, but when used alone it is correctly applied only to the members of the Diptera The forms most commonly known as maggots are actually dipterous larvae and keds (Hippoboscidae) are parasitic forms of flies that have lost their wings The Diptera is the most important group of insects considered in medical entomology They are the vectors [illegible] parasites and diseases *See* ENTOMOLOGY, [illegible] NOMIC, MYIASIS

In number of species it is the [illegible] largest order only the Coleoptera and [illegible] larger At the most recent [illegible] 85 000 kinds of flies were [illegible] the world of which about [illegible] north of Mexico

The order is ubiquitous [illegible] than any other of the insects [illegible] every continent and [illegible] except for the coldest [illegible] arctic and at the [illegible] ranges They occur [illegible] [illegible] niches The [illegible]

of many thousands of feet by specially constructed traps attached to airplanes

TAXONOMY

The present classification of the order and the names of all families in common use today are summarized in the accompanying list The common names are shown for those families of outstanding interest

Orthorrhapha A suborder of the Diptera divided into two series the Nematocera and Brachycera In this group of flies the adult escapes from the puparium through a T shaped opening formed by a longitudinal dorsal split behind the head and a transverse split at the front of this

Nematocera The adults of this series have antennae that are usually longer than the head the flagellum consists of 10 to 65 segments generally of similar size and shape These segments may be surrounded by delicate loops of sensory organs and covered by long hairs especially in males The anal cell of the wing is usually open or absent and the palpi are pendulous

Brachycera Adults have antennae that are usually shorter than the head, the flagellum is reduced to a single segment with or without an arista In some families, there is a distinct indication of segmentation or annulation in the third segment Palpi are porrect and the anal cell is closed or narrowed before the posterior margin of the wing

Cyclorrhapha This suborder is also divided into two series the Aschiza and Schizophora Developing adults are always enclosed in a puparium (the hardened transformed last larval skin), from which they emerge by pushing off the anterior end of the puparium leaving a circular opening

Aschiza The adults of this series lack a frontal suture and ptilinum or frontal sac A small crescent shaped sclerite, the frontal lunnule is absent or indistinct

Schizophora This series includes all the remaining flies usually divided into the sections Myodaria and Pupipara The adults possess a frontal suture through which a distensible sac, or ptilinum is pushed to help the young developing adult escape from its pupal case

Suborder Orthorrhapha
Series I *Nematocera*
- Melusinidae
- *Tanyderidae* (primitive crane flies)
- Tipulidae (crane flies)
- Liriopeidae
- *Dixidae*
- Sylvicolidae
- Psychodidae (moth flies)
- Deuterophlebiidae
- Tendipedidae (midges)
- Heleidae (biting midges)
- Chaoboridae (phantom midges)
- Culicidae (mosquitoes)
- Lycoriidae
- Fungivoridae (fungus gnats)
- Itonididae (gall midges)
- Bibionidae (march flies)
- Scatopsidae (minute black scavenger flies)
- Simuliidae (black flies)
- Blephariceridae
- Thaumaleidae

Series II Brachycera
- Tabanidae (deer and horse flies)
- Pantophthalmidae (wood boring flies)
- Stratiomyidae (soldier flies)
- Xylomyiidae
- Erinnidae
- Coenomyidae
- Rhagionidae (snipe flies)
- Nemestrinidae (hairy flies)
- Acroceridae (hump-backed flies)
- Bombyliidae (bee flies)
- Therevidae (stiletto flies)
- Omphralidae
- Asilidae (robber flies)
- Mydaidae (mydas flies)
- Apioceridae
- Dolichopodidae (long legged flies)
- Empididae (dance flies)
- Musidoridae

Suborder Cyclorrhapha
Series I Aschiza
- Phoridae (hump backed flies)
- Thaumatoxenidae
- Clythiidae (flat footed flies)
- Dorilaidae (big headed flies)
- Syrphidae (flower flies)

Series II Schizophora
Section I Myodaria (muscids)
Subsection I Acalypteratae
- Conopidae (wasp flies)
- Clusiidae
- *Helomyzidae* (sun flies)
- Sphaeroceridae
- Coelopidae
- Dryomyzidae
- Sciomyzidae
- Celyphidae
- Lauxaniidae
- Periscelidae
- Lonchaeidae
- *Pallopteridae*
- Otitidae (otitid flies)
- Tachyniscidae
- *Tephritidae* (fruit flies)
- Pyrgotidae
- *Rhopalomeridae*
- Tanypezidae
- Neriidae
- Micropezidae
- Phytalmiidae
- Sepsidae (spiny legged flies)
- Piophilidae (skipper flies)
- Psilidae (rust flies)
- Diopsidae (stalk eyed flies)
- Canaceidae (seashore flies)
- Ephyridae (shore flies)
- Chloropidae (chloropid flies)
- *Asteidae*
- Drosophilidae (vinegar flies)
- Megamerinidae
- Chryomyidae
- Opomyzidae
- Agromyzidae (leaf miner flies)
- Cryptochaetidae
- Tethinidae
- Odiniidae
- Milichiidae
- Anthomyzidae
- Chamaemyidae (aphid flies)

Subsection II Calypteratae
- Muscidae (house flies stable flies and allies)
- Scopeumatidae (dung flies)
- Sarcophagidae (flesh flies)
- Calliphoridae (blow flies)
- Hypodermatidae (warble flies)
- Larvaevoridae (tachina flies)
- Cuterebridae (rabbit bots rodent bots)
- Oestridae (bot flies)
- Gastrophilidae (horse bots)
- Glossinidae (tsetse flies)

Section II Pupipara
- Hippoboscidae (louse flies)
- Streblidae (bat flies)
- Nycteribiidae (bat tick flies)
- Braulidae (bee lice)

The section Myodaria which includes a large group of families contains half or more of the de scribed species of flies Almost without exception the antenna of the ad ilt consists of three segments *The third segment is without any sign of rings or* annulations and almost always has a dorsal arista All families except the Conopidae have the second cubitus vein united with the second anal vein for nearly their entire lengths

The subsection Acalypteratae of the Myodaria is a large group in which the members of the families are small in size The alulae or calypters are small or rudimentary

Representatives of the Calypteratae subsection have well-developed calypters of moderate size They include all of the larger forms known as flies such as the housefly and stable fly

The section Pupipara contains species which are parasitic in the adult state They have small eyes an indistinctly segmented abdomen and all pairs of legs are widely separated at their attachment to the body Eggs hatch and larvae grow to maturity within the abdomen of the mother

MORPHOLOGY

Adult flies vary from somewhat less than 1 mm to over an inch in body length The head is vertical usually with three ocelli at the dorsal vertex and the mouthparts ventral In some cases however the head is distinctly longer than high when viewed laterally The compound eyes are usually large and prominent and are either holoptic (touching at the top of the head) or dichoptic (separated at the vertex by part of the head capsule)

Mouthparts The mouthparts are modified for either lapping or piercing but never with the man dibles apposable and capable of chewing The structure of the mouthparts varies greatly through out the order so that their homologies are often rather difficult to determine In the mosquitoes and some other Nematocera the labrum mandibles hypopharynx and maxillae are greatly elongated and form a piercing tube enclosed by the labium or labellum Throughout the Orthorrhapha there tends to be a retention of mandibles for piercing *especially in those flies that require blood meals* In the Cyclorrhapha on the other hand mandibles tend to be reduced or lost entirely and with few exceptions the work of the mouthparts is largely sucking with the maxillae and labium taking over the large share of work In these flies the labellum is often enlarged and flattened and provided with minute rasping spines and a specialized mechanism for taking up fluids Since all adult flies imbibe fluids only the presence of a pharyngeal pump to transfer fluids from the external substrate to the gut is characteristic

Antennae Each antenna consists typically of two basal segments and one or more additional seg ments called the flagellum In the Nematocera the flagellum consists of a variable number of quite similar segments In all remaining flies there are three distinct antennal segments The third seg ment representing the flagellum is usually longer than the two basal ones and sometimes complex It is either annulated or produced in various shapes, and often bears an arista which is bare or haired

In the Cyclorrhapha a small crescent shaped sclerite the lunule is situated just above the an tennae The flies in which this sclerite is separated from the rest of the frontal portion by a suture have an internal distensible sac or ptilinum which can be pushed through this suture before the fly is fully developed It is used to break open the end of the puparium to allow the fly to escape Once hard ened and fully adult the fly looses all external trace of this sac

Thorax The prothorax and metathorax are small and closely united with the large mesothorax which contains the musculature for the single pair of wings The legs are usually alike except in some

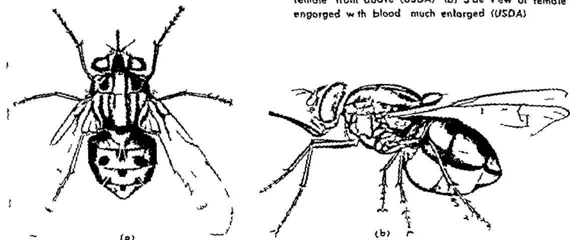


Fig 1 (a) Stable fly *Stomoxys calcitrans* (L) adult female from above (USDA) (b) Side view of female engorged with blood much enlarged (USDA)

species in which the prothoracic pair are raptorial
d ... th ... e spec es n which the fore mid or hind

the c are

present in almost all species Beneath each claw in many species there is a membranous pad, or pulvillus In addition some flies carry an empodium between each pair of pulvillae This may be bristle like or may take the form of a membranous pad like the pulvillae in shape and size

Wings. Only the first pair of wings the mesothoracic pair, are developed The second pair is replaced by a pair of club shaped organs or halteres which serve as balancing organs in flight They are present in most species even when the mesothoracic wings are absent In many species especially those of the Nematocera abundant scalelike setae clothe the veins and sometimes the margins of the wings The wings of some of the higher Diptera bear yellow, brown or black markings that aid in species identification The venation of the primitive dipteran wing follows quite closely that of the primitive insect wing, except that vein R radius, tends to be reduced, and only the important cross veins are present In the higher Diptera the branches of some of the veins coalesce with others thus further reducing the number of apparent veins present In the calypterate flies, a pair of membranous lobes the squamae or calypters are found at the extreme base of the wing Each of these occurs in pairs, the upper half fastened to the wing and the lower half to the thorax In many acalypterate flies these structures are present but greatly reduced

Setae. In certain families of Diptera, some of the setae covering the head and body are greatly enlarged and are used for identification The location and relative size of these bristles are often characteristic of species, genera, or families

Egg, larva, and pupa The eggs are usually elongated and taper at both ends They are relatively

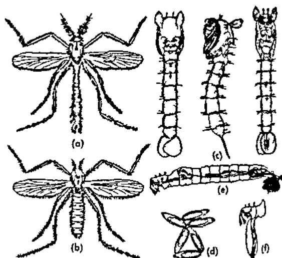
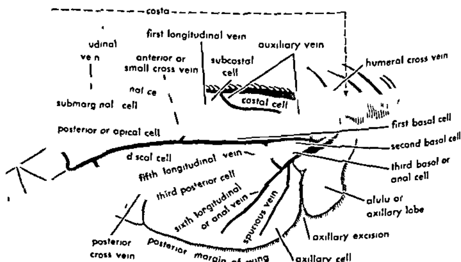


Fig 3 The Clear Lake gnat, *Chaoborus astictopus* Dyar and Shannon (a) Male, (b) female, (c) pupa (d) eggs, (e) larva (phantom larva), (f) eversible pharyngeal sac extended from the mouth (after W Herms 1937) (From E O Essig, College Entomology, Macmillan, 1942)

smooth surfaced, but the integument is often characteristically sculptured in designs which are visible only under high magnification The eggs of some aquatic species are provided with air sacs, of characteristic design, which aid in keeping them afloat When laid, most eggs are whitish but darken gradually as the embryo grows within them

Larvae, unlike adults, have elongate bodies Mouthparts of this feeding stage are usually sclerotized structures They are variously and many times intricately designed for rasping or chewing but sometimes they are reduced or apparently entirely absent The head of the larva of most lower Diptera is large, sclerotized, and complicated in internal structure, while in many species of higher Diptera it is reduced in size The mouthparts are



viridescens Desvoidy
E O Essig, College

always well developed to enable the feeding stage to reach maturity readily Although dipterous larvae never have true segmented appendages anterior or posterior prolegs may be present Nearly all types of spiracular arrangement is found in this order but many larvae have prominent prothoracic and posterior spiracles with characteristic openings

Pupae have appendages which are rather adherent to the body In some pupae the developing adult is free and in others it is encased in a hard ened last larval skin resembling a seed or capsule and presenting no particular external features except spiracles and segmentation

LIFE CYCLE

The adult stage in the life history may be regarded as a reproductive and dispersal phase during which these insects increase both in number and often in geographic range Females select suitable oviposition sites and lay a variable number of eggs after which they usually die The eggs after a period of incubation of varying length give rise to larvae which represent a growth phase During this period the insect does almost nothing but nourish itself thus providing the tissues necessary for a later transformation into an adult This growth period is divided into stages or instars each of which terminates with the molting of the larval skin to allow increase in size in the stage that follows In the Diptera there are four larval instars In most of the Orthorrhapha especially those with aquatic immature forms all four instars are active but in the Cyclorrhapha at the third larval molt a hard puparium is formed inside which a quiescent fourth larval stage occurs At the conclusion of the final larval instar there is a dramatic reorganization in tissue structure initiated by the action of appropriate hormones on the so-called "imaginal buds a process that eventually results in the formation of an adult insect During this time the insect is called a pupa because its external form is now altered and external activity nearly ceases

When the adult tissues are almost fully formed the fly emerges from its pupal case and spends some time in drying hardening and expanding its wings attaining color and reaching sexual maturity It is then ready to start its cycle once again

BIOLOGY

Adults The adults most of which are able to fly represent the reproductive and dispersal phase of a fly s developmental cycle After extricating itself from the puparium it requires some period of exposure to the air after which it flies about to find food and to procreate

Nutrition Food requirements of the females apparently differ from those of males because of their need for nutrient materials for developing and maturing eggs In many of the nematocerous Diptera there is quite a different arrangement of mouth parts in the female and nowhere is this more true than in the mosquitoes of which most females require blood meals before they are able to develop

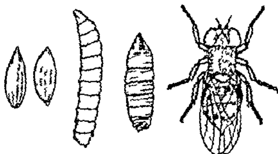


Fig 4 The coachella eye gnat *Hippelates pusio* Loew Eggs larva pupa and adult (From E O Essig College Entomology Macmillan 1942)

eggs Many female mosquitoes have decided preferences for certain hosts Some few species generally regarded as our most important disease vectors are attracted principally to man others prefer a wide variety of mammals and still others birds

The families of flies whose females require blood meals are scattered throughout the order but are concentrated mainly in two areas of the table given above Series I of the Orthorrhapha contains the families Psychodidae Heleidae Culicidae and Simuliidae some or all of whose species have this habit Females of the family Tabanidae in Series II also seek man and animals for blood Among the calypterates the Glossinidae require blood and many other species scattered among various families are attracted by exudates from the bodies of animals and man All four families of the Pupipara apparently require blood meals and live an almost entirely parasitic existence

The adults of a fairly large number of species are predators and feed on all kinds of smaller arthropods This habit is scattered throughout the order to such an extent that a listing of all the families would be too lengthy Whether such food materials are required for the development of their eggs is not known with certainty

Those flies whose larvae live and develop in plant materials have adults with a great variety of food habits It is known that adult Tephritidae of many species are especially attracted to exudates from scale insects which are so numerous on many fruit trees that their honeydew covers large areas of foliage It has been shown that certain components of this material are essential for the development of fruit fly eggs

In the large majority of flies the food habits of adults have never been studied and nothing is known except as the result of an occasional naturalist s observation

Reproduction Most female flies require a period of time between their emergence from the puparium and the time they are ready to lay eggs The length of this period is sometimes characteristic of the species Thus certain fruit flies are able to lay eggs 3 or 4 days after emergence while others require as long as 30 days The females of many species will not accept males for copulation until a certain amount of time has expired some species

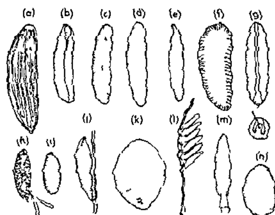


Fig 5 Eggs of Diptera (a) *Hippelates pusio* Loew (after W Herms and A Burgess) (b) *Stomoxys calcitrans* (Linn) (after F Bishopp USDA) (c) *Cochliomyia macellaria* (Fab) (after F Bishopp USDA) (d) *Musca domestica* Linn (after L Howard USDA) (e) *Tabanus punctifer* O S (after Webb and Wells USDA) (f) *Cryptochaetum iceryae* (Williston) (after W Thorpe) (g) *Cochliomyia americana* Cushing and Patton (after E Lake USDA) (h) *Drosophila melanogaster* Meigen (after G Martelli) (i) *Tipula infuscata* Loew (after J Hyslop USDA) (j) *Gasterophilus nasalis* (Linn) (after F Bishopp and W Dove USDA) (k) *Zenillia libatrix* (Panzer) (after Dowden) (l) *Hypoderma lineata* De Villers attached to a hair and (m) dorsal aspect of single egg (USDA) (n) *Simulium simile* Malloch (after A Cameron Canada Dept Agr) (From E O Essig, College Entomology, Macmillan 1942)

require fertilization after every egg laying period and some need fertilization only once during their lifetimes Males usually emerge before the females presumably to assure that they will be sexually mature in time to fertilize the first eggs that are ready for oviposition

Fertilization takes place in many ways throughout the order Males of many of the nematocerous flies and representatives of other families as well form mating swarms These are usually organized over an elevation in the surrounding terrain such as a mound of earth bush tree stump or fence post The swarm remains stationary over the object while [illegible] of such swarms is to accomplish the fertilization of those females.

The females of other species especially many of the Tephritidae and perhaps of many other families are fertilized at about the time they lay their eggs Males of some of these species lie in wait on the surface of a fruit or some other substrate where they apparently know a female will alight to lay eggs The male will attack just before during or immediately after the act of oviposition Males have been seen to pull the ovipositor of a female from the pulp of a fruit in order to deposit the sperm

Eggs. The study of fly eggs has been neglected by biologists and most of their observations have been confined to the various ways in which the females deposit them Almost every conceivable kind of behavior accompanies egg laying Some flies broadcast their small eggs while on the wing spreading them about on the surrounding foliage The eggs of some of these flies are eaten by caterpillars where they hatch into larvae which feed within their hosts Females of species whose larvae are parasitic cement their eggs on the surface of their host or deposit them inside the bodies of the larvae pupae or adults that serve as their victims Some of them are able to deposit their eggs while their hosts are in full flight Perhaps the most interesting habit is that of the female of *Dermatobia hominis* which captures a passing mosquito upon which to fasten her eggs As soon as this carrier has come to rest on an animal to feed the *D hominis* eggs quickly hatch and the larvae enter the mosquito's host

Many flies whose larvae attack plant material have a sharply pointed inflexible ovipositor to insert their eggs within the plant tissues thus affording newly hatched larvae easy access to the succulent parts of the plant Others without such an ovipositor fasten their eggs on the plant surface and the newly hatched larvae immediately burrow into the plant tissue or seek a suitable place to gain entrance The eggs of Sarcophagidae and some other flies hatch within the abdomen of the female and grow to various stages of development before escaping to continue a life of their own In several genera of Itonididae or gall midges the reproductive organs develop prematurely within certain larvae and produce eggs which hatch internally so that only the new or daughter larvae escape from the body of the mother larva

Larva. It has been estimated that the immature stages of about one-half the known species of flies live in association with water Dipterous larvae are found in almost every conceivable habitat within this aquatic environment Some larvae occupy extreme situations such as intertidal wave swept rocks on the seacoasts (Tendipedidae) thermal inland waters as hot as 120°F (Stratiomyidae) saturated brine pools (Ephydridae) and even natural petroleum seeping from the ground (Ephydridae) Adaptations such as these are unusual and occur only in a few species of the families mentioned They are usually accounted for by special and unusual morphological adaptations

Most aquatic larvae live in all kinds of standing water or in running water from the smallest trickles to the large slowly flowing rivers The species that are found as aquatics in the ordinary sense that is swimming under and at the surface are often provided with posterior spiracles especially adapted for obtaining atmospheric oxygen These may be placed at the apex of a more or less elongated "air tube" such as found in many mosquito larvae or on other specially adapted organs They are almost always accompanied by a special arrangement of



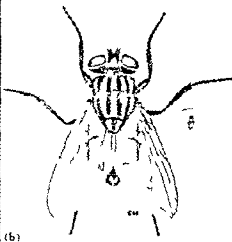


Fig 6 (a) Bot fly larvae attached to the inner lining of the stomach of a horse (Southwestern Biological Supply Company Dallas Texas) (b) Cattle grub adult fly enlarged and natural size (Canada Dept Agr)

hairs and by a water repellent coating so that water will not be inhaled at the surface The larvae of some species may never come to the surface they obtain their oxygen by absorption through membranes in gills or through body surface membranes

The remaining half of the order has larvae that inhabit nonaquatic habitats Of these a very large number of species are found in association with living plants A large share of the Cecidomyiidae cause galls to form in plant tissue or live in galls produced by other species Occasionally galls may be found underground in roots and stems and often in the aerial roots of epiphytes More often they attack the growing stems, leaves, fruits and seeds but no part of a plant can be mentioned that is not subject to a dipterous attack of one kind or another The gall midges and certain tephritids characteristically make a large number of different kinds of galls They are found as inquilines in galls made by other species and may even be found in masses of pitch made to exude from evergreens by their presence

The fruit flies, Tephritidae are so called because of their habit of feeding on ripe or nearly ripe fruit or because they develop in the ovaries and seeds of those plants that do not bear succulent fruit This occurs especially in tropical and subtropical areas of the world Larvae of other species feed directly upon the plant tissue while certain Muscidae attack only the roots The Agromyzidae and certain other small acalypterates feed between layers of the plant tissue in stems and leaves causing mines that are often characteristic of the species making them

Another large group of larvae including those of another part of the Itonididae are found in association with dead and decaying vegetation lying on the ground surface and a number of tephritids and near relatives are strict saprophytes

Many fly larvae are predaceous and are sometimes provided with specially adapted mouthparts for preying upon other species True parasitism occurs here and there throughout the order Larvae of some species of Tachinidae Asilidae Bombyliidae and others are parasitic in the bodies of a large number of other insects Sciomyzidae larvae feed principally on the soft bodies of snails Pyrgotidae on June beetles Cryptochaetidae on scale insects and so forth

Some of the more advanced flies of the calypterate group require the blood or other fluids of mammals for larval development Most of these affect domestic livestock Thus the primary screwworm causes wounds in the undamaged skin of large domestic animals and the *bot* flies have a complicated development during which their larvae travel into many parts of the bodies of their hosts

The Congo floor maggot requires human or other animal blood and commonly lives in association with man feeding through his skin but is not permanently attached Larvae of the Hippoboscidae are carried full grown in the ovaries of the mother but the mother in turn is a permanent resident of sheep and feeds on the sheep s blood

Pupae As stated in the beginning section of this article pupae may be of two kinds In most if not all of the Orthorrhapha all four larval instars are active feeding stages They are either aquatic or phytophagous and the skin of the last larval stage is cast free and lost at the formation of the pupa Some of these pupae especially of the mosquitoes are free swimming just like the larvae while others are quiescent forms In the Cyclorrhapha, the skin of the third larval stage is retained and transformed into a barrel shaped puparium inside which the fourth larval stage is quiescent and in which the pupa reaches full development The various sta of this process have been described in detai the apple maggot *Rhagoletis pomonella*

clorrhaphous pupae are adapted for life under ground or in other well protected parts of their environment and larvae commonly burrow into the soil or these other locations to pupate

No matter what the pupal environment however provision is always made for the escape of the adult In many of the aquatic Diptera pupation takes place at the air water interface mosquito pupae for instance simply float at the water surface until the adult is free Adults of the Simuliidae rise to the water surface from pupae at the bottom surrounded by a bubble of air Pupae of many species are provided with spines or hairs to help them move about

IMPORTANCE

Diptera have probably more economic importance than any other insect order This importance comes from the relatively few species that affect man domestic animals and plants The vast majority of forms have no direct importance although a number are beneficial in one way or another Many flies visit flowers and are efficient plant pollinators Adults may prey upon harmful arthropods or lay eggs in their bodies so that the developing larvae can use them as a source of food Aquatic larvae and pupae under certain circumstances serve as abundant food for fish and other aquatic life while the larvae of a number of terrestrial species act as scavengers assisting bacteria to destroy all kinds of decaying animal and vegetable matter

Human Only a relatively few species of flies cause severe economic loss to man Perhaps of most concern is the role played by flies in disease transmission About 70 species of *Anopheles* mosquitoes transmit an estimated 500 000 cases of malaria each year This disease has serious debilitating effects and occurs on every continent and many islands of the world (*see* MALARIA) Yellow fever once in danger of completely stopping work on the Panama Canal is transmitted principally by a single mosquito *Aedes aegypti* This disease is largely under control yet the same virus has recently been found in monkeys living in tropical rain forests of Central and South America and Africa and is transmitted by *Aedes* and *Haemagogus* species to people who work out of doors in the jungle (*see* YELLOW FEVER) Dengue or breakbone fever is a usually non fatal disease of world wide distribution that leaves its victims debilitated for several weeks It is transmitted by *A aegypti* and *A albopictus* (*see* DENGUE FEVER) Filariasis primarily a disease of peoples of Africa the Orient and the Pacific Islands is caused by a minute roundworm whose larvae are transmitted by a few species of *Anopheles Mansonia Culex* and *Aedes* Filariasis affected a large number of American troops during World War II Long stand ng cases in natives of areas where the d sease is endemic may result in an enlargement of the extremities called elephantiasis (*see* FILARIASIS) With modern methods used by the bacteriologists and virologists a large number of virus d seases known to be transmitted by mosquitoes have been found in man Many of these such as Sindbis virus of Egypt may not produce clinical symptoms of disease but others such as the equine encephalitides which include western eastern St. Louis Japanese and others may be quickly fatal in man With the exception of the viruses none of the diseases mentioned above are transmitted by any insects other than members of the order Diptera. *See* HAEMOSPORIDIIDA

A few less well known diseases are transmitted by other flies The black flies transmit *Onchocerca volvulus* the causative agent of onchocerciasis a disease affecting the eyes of natives of Central and parts of South America (*see* FILARIOIDEA) Sand flies of the genus *Phlebotomus* transmit the organisms that cause kala azar Oriental sore pappataci fever and Oroya fever (*see* LEISHMANIASIS) Even today large parts of Africa remain underdeveloped because of the presence of sleeping sickness a fatal disease transmitted by tsetse flies Species of the genus *Chrysops* (Tabanidae) are instrumental in carrying tularemia primarily a disease of rodent

Domestic animals Domestic animals hides, meat and dairy products are affected by disease transmission or by direct attack by flies Anthrax tularemia botulism many virus diseases and nagana a form of sleeping sickness are some of the diseases transmitted by members of the Diptera that take an annual toll of millions of dollars

Some flies wreak their damage by direct attack The primary screwworm fly deposits eggs on the hides of animals and the larvae upon hatching burrow through the skin and into the flesh The secondary screwworm gains entrance through holes often infected already present on the skin surface *Hypoderma lineata* and *H bovis* are bot flies of special importance The eggs of both species are laid on hairs of cattle and the hatching larvae bore through the skin into the connective tissue During their development they wander through the tissues of the animal and when mature they escape through holes which they make along the spine of their host and drop to the ground to pupate Horses are afflicted by a species of *Gasterophilus*

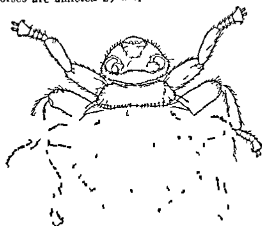


Fig 7 The bee louse *Braula coeca* Nitzsch (From E O Essig *College Entomology* Macmillan 1942

They lick the eggs from their bodies (Fig 6a) and the larvae settle and dwell in their stomachs, often in large numbers Sheep are victims of the sheep bot fly whose larvae live in their nasal passages and sinuses All of these insects affect the meat milk and wool production of animals through debilitation and irritation and hides may be rendered completely useless by their entrance and exit holes An interesting dipteran is the bee louse which is a wingless ectoparasite of the honeybee (Fig 7)

Many species of calypterate flies are attracted by the unsanitary conditions about garbage cans and rendering plants by undressed wounds and by seepages from infected eyes and mucous membranes of living animals By coming into contact with these products flies can mechanically carry on their feet and body hairs the organisms that cause typhoid dysentery diarrhea cholera yaws and trachoma in addition to certain parasitic worms

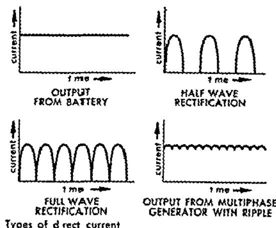
Plants Economic losses as a result of damage by flies to crops are perhaps not as great as those caused by other insects yet their presence has necessitated the expenditure of large sums of money for control Among these perhaps the mo t important are members of the family Tephritidae the larvae of which feed upon and ruin the succulent flesh and seeds of their hosts The most important of these are the European and American cherry fruit flies, their relatives in the genus *Rhagoletis* that attack walnuts in North America six or seven species of the genus *Anastrepha* which attack many kinds of fruit in the New World the Mediterranean fruit fly which is a limiting factor in the production of many fruits especially citrus in many parts of Africa Central and South America and the Mediterranean Basin and several species of the genus *Dacus* which attack olives citrus fruits many kinds of vegetables, and other edible plants and plant parts in Europe the Orient and Pacific Islands

Larvae of some flies mine the leaves of ornamentals thereby defacing them reducing their growth potentials and affecting their production and sale by nurseries The Hessian fly an itonidid is a serious pest of wheat and other grains grown for human consumption in Europe northern Asia and North America and has caused untold losses to growers in those areas Several other species of Itonididae have affected the growth of rice and other plants grown for human consumption

The larvae of some species of Muscidae are known as root maggots and burrow into the underground parts of plants with considerable loss to growers of truck crops over the world [R H F]

Direct current

Electric current which flows in one direction only as opposed to alternating current The current may be of constant magnitude (as when produced by a battery) or it may vary with time (as rectified alternating current or the output from a single pole dc generator) The fluctuation of generated direct current is called ripple In most applications the complete absence of ripple is not essential



Types of direct current

In parts of Europe direct current is still extensively used commercially whereas in the United States it has largely been replaced by alternating current Direct current cannot readily be changed from one voltage to another and so cannot economically be transmitted long distances over cross country power lines *See* ALTERNATING CURRENT CURRENT ELECTRIC [J W ST]

Direct-coupled amplifier

There are many different situations where it is necessary to amplify signals that have a frequency spectrum which extends to zero frequency Some typical examples are amplifiers in electronic differential analyzers (analog computers) certain types of feedback control systems and some medical instruments such as the electrocardiograph Amplifiers which have capacitor coupling between stages are not usable in these cases because the gain at zero frequency is zero Therefore a special form of amplifier called a dc amplifier is necessary These amplifiers will also amplify ac signals *See* AMPLIFIER

Direct coupled dc amplifiers The coupling capacitor causes the gain of a resistance-capacitance (*RC*) coupled amplifier to become zero at zero frequency However some type of coupling circuit must be used between successive amplifier stages to prevent the relatively large plate supply voltage of one stage from appearing at the grid of the following stage These circuits must pass dc signals with the least possible amount of attenuation

One method of coupling is by means of gas tubes such as neon tubes or voltage regulator tubes This method in its simplest form is illustrated in Fig 1 The neon tube and the voltage-regulator tube possess the property common to gas tubes that once ionization has taken place there is a range of values of current through the tube for which the voltage across the tube is nearly constant In effect the gas tube acts as a battery The nearly constant voltage drop E_T is subtracted from E_b to give E_o For example if E_b is +150 volts and E_o is 1 volt, E_T = 149 volts If E_b changes by ΔE_b E_o also changes by ΔE_b If the 149 volt drop were obtain

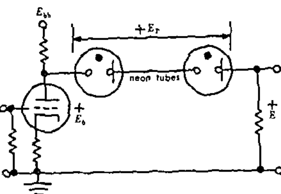


Fig 1 Dc amplifier with gas-tube coupling

by using a resistor then ΔE_o would equal 1/150 of ΔF_b and the stage may have a gain of less than unity

The use of gas tubes as coupling elements can produce good amplifiers in situations where the noise introduced by the gas tube can be tolerated because the input signal is sufficiently large Much effort has been directed toward producing better dc amplifiers and as a result there are amplifiers available which can reliably amplify signals as small as a few microvolts

The large attenuation introduced by the coupling resistor in place of the gas tubes can be greatly reduced by returning the grid leak resistor to a negative supply voltage instead of ground This variation greatly improves the gain producing a ΔE on the order of $\frac{1}{2}\Delta E_b$ (instead of $\frac{1}{150}\Delta E_b$) The additional requirement of a negative voltage supply is easily met the added cost being the only drawback

A circuit which has wide application in dc amplifiers is illustrated in Fig 2 It is a cathode follower feeding a grounded grid amplifier Each tube is half of a twin triode The negative feedback introduced by the cathode resistor aids in stabilizing the performance against undesirable effects

Dc amplifier performance can suffer from effects which are not observed in an *RC* coupled amplifier, because the effects occur at frequencies far below the lower cutoff frequency One effect is cathode drift caused by changes in the cathode emission In the cathode coupled circuit a change in the emission of one tube is equivalent in its effect to changing the input signal e_s A second undesirable effect is the changing of the operating point because of a change in the tube characteristics caused by aging warming up or even vibration of the tube The changes are called residual drift Since cathode drift and residual drift produce changes in the output voltage e_o which are indistinguishable from changes in the signal voltage it is apparent that these undesirable effects must be reduced to a minimum through the use of aged and reliable tubes and components and by sophisticated circuitry

Carrier dc amplifier. A method of amplifying dc (or slowly varying) signals by means of ac amplifiers is to modulate a carrier signal by the signal to be amplified, amplifying the modulated signal and demodulating at the output (In some applications such as instrument servomechanisms, output in the form of a modulated carrier is required and no demodulation is necessary) One arrangement is illustrated in block diagram form in Fig 3

An analysis of an actual circuit would show that in order to have an output e_o free from harmonics introduced by the modulation, the output low pass filter cutoff frequency must be small compared to the modulation frequency This limits the band width of the input signal e_s to a small fraction of the bandwidth of signals which can be amplified by the types previously discussed This is a disadvantage in most cases but there are cases such as the output voltage from a thermocouple, where the required bandwidth is small Various types of choppers many of them electromechanical in nature are used as modulators (*see* VIBRATOR) An additional disadvantage is that the chopper must be carefully designed to reduce to a minimum the hum and noise which can be introduced However this amplifier provides stable amplification of the input signal and it is used in many industrial recording instruments *See* MODULATION, VOLTAGE AMPLIFIER

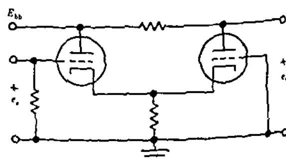


Fig 2 Dc cathode-coupled amplifier

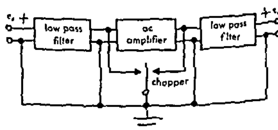
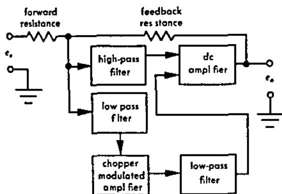


Fig 3 Carrier dc amplifier

Chopper stabilized a[illegible] **ers** A chopper stabilized amplifier [illegible] two ar[illegible] one *amplifier* is a [illegible] fie[illegible] ng a chopper to m[illegible] n of the signal to [illegible] r is a straight dc [illegible] nt in reducin[illegible] lt appear [illegible] *the d*[illegible] bloc[illegible] illus[illegible]



Fg 4 Chopper stab lized dc ampl f er

A component of the output signal e_o is due to drift in the dc amplifier and if it is fed back through the feedback resistor to the input of the chopper modulated amplifier (it cannot be fed back to the dc amplifier because of the high pass filter) the drift component will be amplified and returned to the dc amplifier An analysis of the circuit shows that the input voltage e_i to the dc amplifier is altered by a factor equal and opposite to the equivalent drift voltage appearing at the input The net result is that the drift can be reduced to a negligible value

A signal appearing at the input terminals is amplified partly in the dc amplifier and partly in the chopper modulated amplifier because the high pass and low pass filters direct the frequency components of the input signal into the separate amplifiers Thus for dc and very low frequency signals the gain is equal to the gain of the ac amplifier times the gain of the dc amplifier while for higher frequency signals the gain is that of the dc amplifier alone However if the gains are large then when the feedback network is connected the gain is equal to a good approximation to the feedback impedance divided by the impedance in the forward path The gain is therefore essentially independent of the open loop gain

Chopper stabilized amplifiers find wide application in analog computers where they are used in integrating and summing amplifiers In many problems solved on analog computers the time scale is reduced This imposes strict requirements on the freedom from drift of the dc amplifiers These requirements are met with amplifiers where a typical value for the dc amplifier gain is 100 000 and a typical value for the gain of the chopper modulated amplifier is 1 000 [H F K]

Bibliography J E Gibson and F B Tuteur *Control System Components* 1958 J D Ryder *Engineering Electronics* 1957

Direct current circuit theory

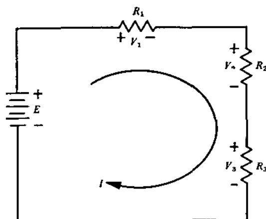
Any combination of direct current (dc) voltage or current sources such as generators and batteries in conjunction with transmission lines resistors inductors capacitors and power converters such as motors is termed a dc circuit Historically the dc circuit was the first to be studied and analyzed mathematically *See* CIRCUIT ELECTRIC

Classification Circuits may be identified and classified into simple series and parallel circuits More complicated circuits may be developed as combinations of these basic circuits

Series circuit A series circuit is illustrated in Fig 1 It consists of a battery of emf E volts and three resistors The conventional current flows from the positive battery terminal through the external circuit and back to the negative battery terminal It passes through each resistor in turn therefore the resistors are said to be in series with the battery

Parallel circuit The parallel circuit shown in Fig 2 consists of a battery paralleled by three resistors In this case the current leaving the positive terminal of the battery splits into three components one component flowing through each resistor then recombining into the original current and returning to the negative terminal of the battery

Physical laws of circuit analysis The operation of the basic series and parallel circuits must obey certain fundamental laws of physics These laws are referred to as Ohm s law and Kirchhoff s laws in honor of their originators



Fg 1 S mple ser es c rcuit

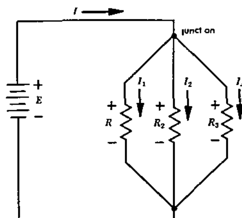


Fig 2 S mple parallel c rcuit

Voltage drops When an electric current flows through a resistor a voltage drop appears across the resistor the polarity being such that the voltage is positive at the end where the conventional current enters the resistor This voltage drop is directly proportional to the product of the *current in amperes* and the resistance in ohms This is Ohm s law Expressed mathematically

$$V = IR$$

Thus in Fig 1 the drop across R_1 is V_1 and has the polarity shown *See* OHM S LAW

Summation of voltages The algebraic sum of all voltage sources (rises) and voltage drops must add up to zero around any closed path in any circuit *This is* Kirchhoff s first law Referring to Fig 1 *the sum of the* voltages about this closed circuit would be

$$E - V_1 - V_2 - V_3 = 0$$

or

$$E = V_1 + V_2 + V_3$$

where the minus signs indicate a voltage drop *Written in terms* of current and resistance this becomes

$$E = I(R_1 + R_2 + R_3) = IR_{total}$$

From this results the important *conclusion* that resistors in series may be added to obtain the equivalent *total* resistance

$$R_{eq} = R_1 + R_2 + R_3 +$$

S • V [illegible]

[illegible]

[illegible] eferring to Fig 2 the current flowing into the junction is I *amperes* while that flowing out is the sum of I_1 plus I_2 plus I_3 Therefore

$$I = I_1 + I_2 + I_3$$

This is Kirchhoff's second law *In this case the* same voltage appears across each resistor Expressed in terms of this voltage and the values of the individual resistors by means of Ohm s law this becomes

$$I = \frac{E}{R_1} + \frac{E}{R_2} + \frac{E}{R_3} = \frac{E}{R_{eq}}$$

The equivalent resistance R_{eq} that can replace the resistors in parallel can be obtained from

$$R_{eq} = \left[\frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \quad\right]^{-1}$$

Therefore resistors in parallel are added by computing the corresponding conductances (reciprocal of resistance) and adding to obtain the equivalent conductance The reciprocal of the equivalent conductance is the equivalent resistance of the parallel combination *See* CONDUCTANCE

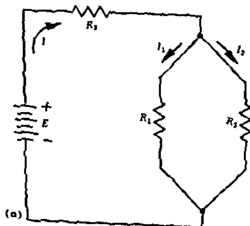
Sources Sources such as batteries and generators may be connected in series and parallel Series connections serve to increase the voltage the net voltage is the algebraic sum of the individual source voltages

Sources in parallel provide the practical function of increasing the net current rating over the rating of the individual sources the net current rating is the sum of the individual current ratings

Series parallel circuits More complicated circuits are nothing more than combinations of simple series and parallel circuits as illustrated in Figs *3a* and *3b*

Single source Circuits that contain only a single source are readily reduced to a simple series circuit In the circuit of Fig *3a* the parallel combination of R_1 and R_2 is computed and used to replace the parallel combination The resultant circuit is now a simple series circuit consisting of R_3 and R_{eq} and can readily be solved for the series current if the voltage is known

Multiple sources Circuits that contain two or more sources located in various branches cannot be reduced to a simple series circuit (see Fig *3b*) The three basic laws of circuit theory still hold and may be directly applied to provide a simultaneous solution to the loop currents i_1 and i_2 flowing in each basic series circuit present in the over all network



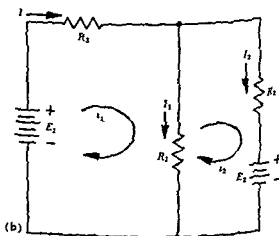


Fig 3 Series-parallel circuits (a) Single source (b) Multiple sources

In this example the summation of voltages around the individual series circuits or loops is

$$E_1 = (R_1 + R_3)i_1 - R_1 i_2$$
$$-E_2 = -R_1 i_1 + (R_1 + R_2)i_2$$

which may be solved for the mathematical loop currents i_1 and i_2 These loop currents can in turn be identified by reference to the circuit where it is seen that i_1 is identical to I and i_2 is I_2 therefore $(i_1 - i_2)$ is I_1

This method may be used to solve any complicated combination of simple circuits Other methods are also available to the circuit analyst *See* NETWORK THEORY ELECTRICAL

Power The electric power converted to heat in any resistor is equal to the product of the voltage drop across the resistor times the current through the resistor

$$P = VI$$

By means of Ohm s law this may also be written as

$$P = VI = V^2/R = I^2R$$

The *total power dissipated in a circuit is the arithmetic* sum of the power dissipated in each resistor

Circuit response In the circuits mentioned thus far the circuit responds in an identical manner from the moment the circuit is excited (switches closed) through any extended period of time This is not true of circuits typified by those of Figs 4 and 5

For instance when the switch of Fig 4 is first closed a momentary current limited only by the resistor R flows As time passes the capacitor C charges and the voltage across it increases eventually reaching a value equal to the applied voltage at which time all flow of current ceases The circuit current is given by

$$i = \frac{E}{R}\epsilon^{-t/RC}$$

amperes The product RC is known as the time constant of the circuit (*see* TIME CONSTANT) The energy W in joules stored in a capacitor at any time is

$$W = \tfrac{1}{2} CE^2$$

For the circuit of Fig 5 the initial current upon closing the switch is zero since any attempt to cause a rate of change of current through the coil L induces a counter emf across the coil or inductor Eventually this counter emf disappears and a steady state current E/R flows indefinitely in the circuit

The current at any time after closing the switch is

$$i = \frac{E}{R}(1 - \epsilon^{-Rt/L})$$

amperes The factor R/L is the time constant of the circuit The energy W stored in an inductance at any time is

$$W = \tfrac{1}{2} LI^2$$

For a complete discussion of transient phenomena *see* TRANSIENT ELECTRIC [R L R]

Bibliography W Timbie V Bush and B G Hoadley *Principles of Electrical Engineering* 4th ed 1951

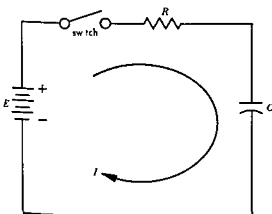


Fig 4 A series *RC* circuit

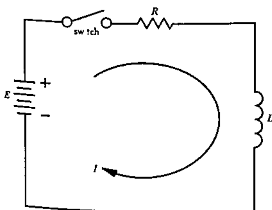


Fig 5 A series *RL* circuit

Direct current generator

A rotating electric machine which delivers a unidirectional voltage and current An armature winding mounted on the rotor supplies the electric power output One or more field windings mounted on the stator establish the magnetic flux in the air gap A voltage is induced in the armature coils as a result of the relative motion between the coils and the air gap flux Faraday s law states that the voltage induced is determined by the time rate of change of flux linkages with the winding Since these induced voltages are alternating a means of rectification is necessary to deliver direct current at the generator terminals Rectification is accomplished by a commutator mounted on the rotor shaft *See* COMMUTATION WINDINGS (ELECTRIC MACHINERY) *See also* ELECTRIC ROTATING MACHINERY GENERATOR ELECTRIC

Carbon brushes insulated from the machine frame and secured in brush holders transfer the armature current from the rotating commutator to the external circuit Brushes are held against

square inch Armature current passes from the brush to brush holder through a flexible copper lead In multipolar machines all positive brush studs are connected together as are all negative studs to form the positive and negative generator terminals In most dc generators the number of brush studs is the same as the number of main poles In modern machines brushes are located in the neutral position where the voltage induced in a short circuited coil by the main pole flux is zero The brushes continuously pick up a fixed instantaneous value of the voltage generated in the armature winding See COMMUTATOR

The generated voltage is dependent upon speed n in rpm number of poles p flux per pole Φ in webers number of armature conductors z and the number of armature paths a The equation for the average voltage generated is

$$E_g = \frac{nz\Phi_-}{60a} \quad \text{volts}$$

The field windings of dc generators require a direct current to produce a magnetomotive force (mmf) and establish a magnetic flux path across the air gap and through the armature Generators are classified as series shunt compound or separately excited according to the manner of supplying the field excitation current In the separately excited generator the field winding is connected to an independent external source Using the armature as a source of supply for the field current dc generators are also capable of self excitation Residual magnetism in the field poles is necessary for self excitation Series shunt and compound wound generators are self excited and each produces different voltage characteristics

When operated under load the terminal voltage changes with change of load because of armature resistance drop change in field current and armature reaction Interpoles and compensating or pole face windings are employed in modern generators to improve commutation and to compensate for armature reaction See ARMATURE REACTION

Series generator The armature winding and field winding of this generator are connected in series as shown in Fig 1 Terminals T_1 and T_2 are connected to the external load The field mmf aids the residual magnetism in the poles permitting the generator to build up voltage The field winding is wound on the pole core with a comparatively few turns of wire of large cross section capable of carrying rated load current The magnetic flux and consequently the generated emf and terminal voltage increase with increasing load current Figure 4 shows the external characteristic or variation of terminal voltage with load current at constant speed Series generators are suitable for special purposes only such as a booster in a constant voltage system and are therefore seldom used

Shunt generator The field winding of a shunt generator is connected in parallel with the armature winding, as shown in Fig 2 The armature supplies both load current I_l and field current I The field current is 1-5% of the rated armature current I_a the higher value applying to small machines The field winding resistance is fairly high since the field consists of many turns of small cross section wire For voltage build up the total field circuit resistance must be below a critical value above this value the generator voltage cannot build up The no-load voltage to which the generator builds up is varied by means of a rheostat in the field circuit The external voltage characteristic Fig 4 shows a reduction of voltage with increases in load current but voltage regulation is fairly good in large generators The output voltage may be kept constant for varying load current conditions by manual or automatic control of the rheostat in the field circuit A shunt generator will not maintain a large current in a short circuit in the external circuit since the field current at short circuit is zero

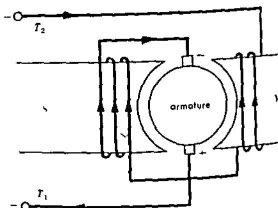


Fig 1 Series generator

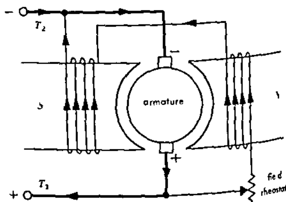


Fig 2 Shunt generator

The shunt generator is suitable for fairly constant voltage applications such as an exciter for ac generator fields battery charging and for electrolytic work requiring low voltage and high-current capacity Modern automobiles use a shunt generator in conjunction with automatic regulating devices to charge the battery and supply power to the electrical system Shunt wound generators are well adapted to stable operation in parallel

Compound generator. This generator has both a series field winding and a shunt field winding. Both windings are on the main poles with the series winding on the outside. The shunt winding furnishes the major part of the mmf. The series winding produces a variable mmf dependent upon the load current and offers a means of compensating for voltage drop. Figure 3 shows a cumulative compound connection with series and shunt fields aiding. A diverter resistance across the series field is used to adjust the series field mmf and vary the degree of compounding. By proper adjustment a nearly flat output voltage characteristic is possible. Cumulative compound generators are overcompounded, flat compounded, or undercompounded as shown by the external characteristics in Fig. 4. The shunt winding is connected across the armature (short shunt connection) or across the output terminals (long shunt connection). Figure 3 shows the long shunt connection.

Voltage is controllable over a limited range by a rheostat in the shunt field circuit. Compound generators are used for applications requiring constant voltage such as lighting and motor loads. Generators used for this service are rated at 125 or 250 volts and are flat or overcompounded to give a regulation of about 2%. An important application is in steel mills which have a large dc motor load. Cumulative compound generators are capable of stable operation in parallel provided the series fields are connected in parallel by an equalizer bus.

In the differentially compounded generator the series field is connected to oppose the shunt field mmf. Increasing load current causes a large voltage drop due to the demagnetizing effect of the series field. This generator has only a few applications, such as arc welding generators and special generators for electrically operated shovels.

Separately excited generator. The field winding of this type generator is connected to an independent dc source. The field winding is similar to that in the shunt generator. Separately excited generators are among the most common of dc generators, for they permit stable operation over a very wide range of output voltages. The slightly drooping voltage characteristic, Fig. 4, may be corrected by rheostatic control in the field circuit. Applications are found in special regulating sets, such as the Ward Leonard system, and in laboratory and commercial test sets.

Special types. Besides the common dc generators discussed in this article, a number of special types may be found in the bibliographical references. These include the homopolar, third brush, diverter pole, and Rosenberg generators. For discussion of the Amplidyne, Regulex, and Rototrol see DIRECT CURRENT MOTOR.

Commutator ripple. The voltage at the brushes of dc generators is not absolutely constant. A slight high frequency variation exists, which is superimposed upon the average voltage output. This is called commutator ripple and is caused by a cyclic change in the number of commutator bars contacting the brushes as the machine rotates. The ripple decreases as the number of commutator bars is increased and is usually ignored. In servomechanisms employing a dc tachometer for velocity feedback, the ripple frequency is kept as high as possible. [R T W]

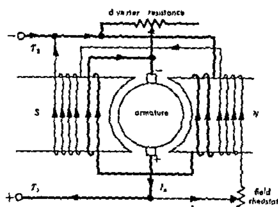


Fig. 3. Cumulative compound generator, long shunt connection.

overcompound — separately excited
flat compound — shunt
undercompound — differential compound
series

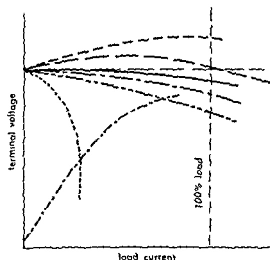


Fig. 4. External characteristics of direct-current generators.

Bibliography. A. E. Fitzgerald and C. Kingsley, *Electric Machinery*, 1952; A. E. Knowlton (ed.), *Standard Handbook for Electrical Engineers*, 9th ed., 1957; A. S. Langsdorf, *Principles of Direct current Machines*, 5th ed., 1940; M. Liwschitz

Direct-current motor

An electric rotating machine energized by direct current and used to convert electric energy to mechanical energy. It is characterized by its relative ease of speed control and in the case of the series connected motor by an ability to produce large torque under load without taking excessive current. Output of this motor is given in horsepower the unit of mechanical power. Normal full load values of voltage current and speed are generally given.

Direct current motors are manufactured in several horsepower rating classifications (1) subfractional approximately 1 35 millihorsepower (2) fractional 1/40 1 horsepower and (3) integral 1 to several hundred horsepower.

The standard line voltages applied to dc motors are 6 12 27 32 115 230 and 550 volts. Occasionally they reach higher values.

Normal full load speeds are 850 1140 1725 and 3450 rpm. Variable speed motors may have limiting rpm values stated.

Protection of the motor is afforded by several types of enclosures such as splash proof drip proof dust explosion proof, dust ignition proof and immersion proof enclosures. Some motors are totally enclosed.

The principal parts of a dc motor are the frame the armature the field poles and windings and the commutator and brush assemblies. The frame consists of a steel yoke of open cylindrical shape mounted on a base. Salient field poles of sheet steel laminations are fastened to the inside of the yoke. Field windings placed on the field poles are interconnected to form the complete field winding circuit. The armature consists of a cylindrical core of sheet steel disks punched with peripheral slots air ducts and shaft hole. These punchings are aligned on a steel shaft on which is also mounted the commutator. The commutator, made of hard drawn copper segments is insulated from the shaft. Segments are insulated from each other by mica. Stationary carbon brushes in brush holders make contact with commutator segments. Copper conductors placed in the insulated armature slots are interconnected to form a reentrant lap or wave style of winding. For a discussion of commutation and windings *see* COMMUTATION, WINDINGS (ELECTRIC MACHINERY).

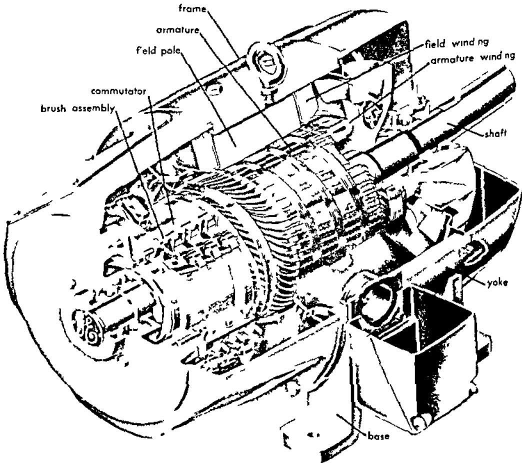


Fig 1 Cutaway view of typical dc motor (General Electric)

DIRECT-CURRENT MOTOR PRINCIPLES

Rotation of a dc motor is produced by an electromagnetic force exerted upon current carrying conductors in a magnetic field For basic principles of motor action *see* MOTOR ELECTRIC

In Fig 2 forces act on conductors on the left path of the armature to produce clockwise rotation Those conductors on the right path whose current direction is reversed also will have forces to produce clockwise rotation The action of the commutator allows the current direction to be reversed as a conductor passes a brush

The net force from all conductors acting over an average radial length to the shaft center produces a torque T given by the expression

$$T = K_t \Phi I_a \qquad (1)$$

where K_t is a conversion and machine constant Φ is net flux per pole and I_a is the total armature current

The voltage E induced as a counter electromotive force (emf) by generator action in the parallel paths of the armature plus the voltage drop I_aR_a through the armature due to armature current I_a and armature resistance R_a must be overcome by the total impressed voltage V from the line The voltage relations can be expressed by the equation

$$V = E + I_aR_a \qquad (2)$$

The counter emf and motor speed n are related by the expression

$$n = \frac{E}{K\Phi} \qquad (3)$$

where K is a conversion and machine constant

Mechanical power output can be expressed by

$$\text{HP} = \frac{2\pi nT}{33\,000} \quad \text{horsepower} \qquad (4)$$

where n is the motor speed in rpm and T is the torque developed in pound feet

By use of these four equations the steady state operation of the dc motor may be determined

TYPES OF DC MOTORS

Shunt motor The field circuit and the armature circuit of a dc shunt motor are connected in parallel (Fig 3*a*) The field windings consist of many turns of fine wire The entire field resistance including a series connected field rheostat is relatively large The field current and pole flux are essentially constant and independent of the armature requirements The torque is therefore essentially proportional to the armature current

In operation an increased motor torque will be produced by a nearly equal increase in armature current (Eq 1) since K_t and Φ are constant Increased I_a produces an increase in the small voltage I_aR_a (Eq 2) Since V is constant, E must decrease by the same small amount resulting in a small decrease in speed n (Eq 3) The speed load curve is practically flat resulting in the term "constant speed" for the shunt motor Typical characteristics are shown in Fig 3*b*

Typical applications are for load conditions of fairly constant speed such as machine tools blowers centrifugal pumps fans conveyors wood and metal working machines steel paper and cement mills and coal or coke plant drives

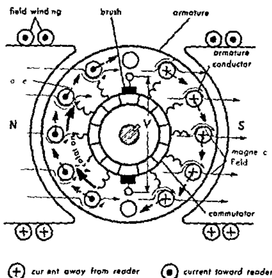
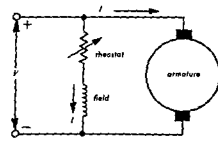


Fig 2 Rotation in a dc motor



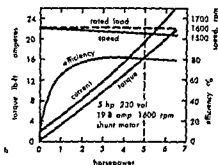


Fig 3 Shunt motor (a) Connections (b) operating characteristics

A stabilized shunt motor is one having a light series winding cumulatively connected (series field aids shunt field) to prevent a rise in speed with load increase Although the pole flux is weakened due to armature reaction the increased armature current causes added flux from the series field The ratio of flux to counter emf (Eq 3) is therefore held more nearly constant than in an unstabilized shunt motor and a more constant speed is maintained The fields may also be differentially connected (series field opposes shunt field) to produce a slight speed decrease with load increase

Series motor The field circuit and the armature circuit of a dc series motor are connected in series (Fig 4*a*) The field winding has relatively few turns per pole The wire must be large enough to carry the armature current The flux Φ of a series motor is nearly proportional to the armature current I_a which produces it Therefore the torque (Eq 1) of a series motor is proportional to the square of the armature current neglecting the effects of core saturation and armature reaction An increase in torque may be produced by a relatively small increase in armature current

In operation the increased armature current which produces increased torque also produces increased flux Therefore speed must decrease to produce the required counter emf to satisfy Eqs 1 and 3 This produces a variable speed characteristic At light loads the flux is weak because of the small value of armature current and the speed may be excessive For this reason series motors are generally connected permanently to their loads through gearing

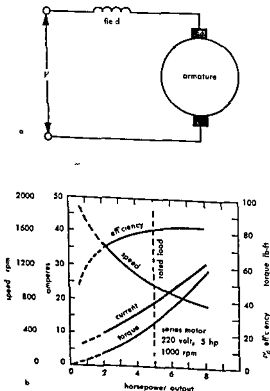


Fig 4 Series motor (*a*) Connection (*b*) Typical operating characteristics.

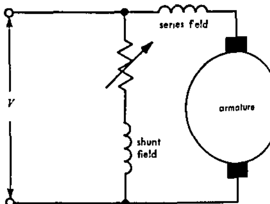


Fig 5 Connection of a compound motor

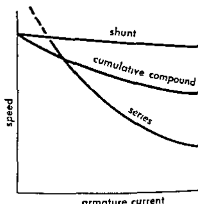


Fig 6 Comparative speed current curves of dc motors

The characteristics of the series motor are shown in Fig 4*b* Typical applications of this motor are to loads requiring high starting torques and variable speeds, for example, cranes, hoists gates, bridges, car dumpers, traction drives and automobile starters

Compound motor. A compound motor has two separate field windings One, generally the predominant field is connected in parallel with the armature circuit the other is connected in series with the armature circuit (Fig 5)

The field windings may be connected in long or short shunt without radically changing the operation of the motor (*see* DIRECT CURRENT GENERATOR) They may also be cumulative or differential in compounding action With both field windings this motor combines the effects of the shunt and series types to an extent dependent upon the degree of compounding In Fig 6 its typical speed characteristics are compared with those of the shunt and series types Applications of this motor are to loads requiring high starting torques and somewhat variable speeds such as pulsating loads shears bending rolls plunger pumps, conveyors elevators, and crushers

Separately excited motor. The field winding of this motor is energized from a source different

from that of the armature winding The field winding may be of either the shunt or series type and adjustment of the applied voltage sources produces a wide range of speed and torque characteristics

Small dc motors may have permanent magnet fields with armature excitation only Such motors are used with fans blowers rapid transfer switches electromechanical activators and programming devices

STARTING AND SPEED CONTROL

Starting DC motors are usually started with a rheostat in series with the armature circuit This motor starting resistor is of the proper rating in watts and ohms to withstand starting currents

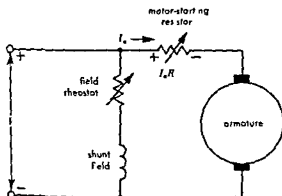


Fig 7 Connection for starting a dc shunt motor

When a dc motor is started the field winding is fully excited Since there is no rotation of the armature no counter emf is generated Therefore the armature current would be dangerously high unless an additional starting resistance were placed in the armature circuit (Eq 2) This rheostat is manually or automatically cut out of the circuit as the motor approaches full speed Small motors with low armature inertia reach full speed rapidly and do not require starting resistors Separately excited motors may be started by control of the voltages applied to the armature

Speed control Speed of a dc motor may be controlled by changing the flux or counter emf of the motor (Eq 3) Adjustment of the armature voltage I will affect the counter emf E (Eq 2) by approximately the same amount The speed n is affected by the change in counter emf according to Eq 3 Insertion of a resistor in the armature circuit would also affect the speed but is seldom used because of the large power losses in the resistor Speed control by adjustment of the applied armature voltage is used extensively where separate adjustable voltage sources are available

A change in flux Φ will also affect speed n (Eq 3) Flux may be changed by a variable resistor in series with the shunt field of a shunt or compound motor This field rheostat should have a total resistance comparable to that of the shunt field and be of sufficient capacity to withstand the relatively small shunt field current

Ward Leonard speed control system In this system the armature voltage of a separately excited dc motor is controlled by a motor generator set A typical circuit (Fig. 8) shows a prime mover M_1 often a three-phase induction motor mechanically coupled to a dc generator G and to an exciter generator E The latter provides field excitation for the dc machines Control of the generator field rheostat R_1 affects the output voltage of the generator G This voltage may be smoothly varied from a low value to a value above normal When this voltage is applied to the armature of the motor M_2 the speed of this motor will be variable over a wide range Additional speed control of motor M_2 may be gained by adjustment of rheostat R_2

The disadvantages of this system are the added equipment and maintenance costs it entails However the wide range and fineness of control in a low current circuit make it applicable to high speed passenger elevators large hoists power shovels steel mill rolls drives in paper or textile mills and the propulsion of small ships

Amplidyne The Amplidyne (dynamoelectric amplifier) is a rotating two-stage power amplifier in which a small change in field power in a dc generator results in a large change in output armature power A large motor connected to the output of the generator may be controlled in speed by adjustment of the relatively small field power of the Amplidyne

In Fig 9 the control field current I_f produces an mmf F_1 The resultant flux and the short circuit of brushes aa' cause the induced voltage e_1 to force a large current i_1 through armature circuit Because of the magnetic core design current i_1 produces mmf F_2 and its resultant flux which induces voltage e between brushes bb Motor M connected across brushes bb will draw a current i_2 which produces an mmf F_3 tending to weaken the original mmf F_1 However compensating windings C energized by i_2 will produce an mmf to oppose F_3 and restore the value of F_1

This dynamic amplifier may produce amplifications of 10 000 to 1 or higher It is applied to a variety of servomotors to control starting acceleration and deceleration Other typical applications include voltage regulation of large ac gen-

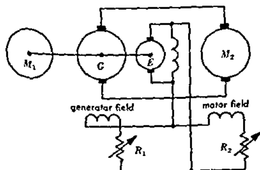


Fig 8 Ward Leonard speed control system

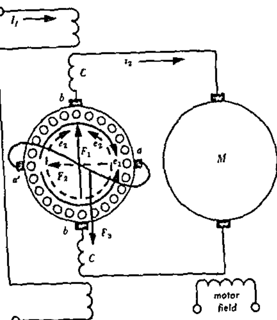


Fig 9 Operation of the Amplidyne

erators dc voltage control in cold strip mills speed control of paper mills positioning control of gun turrets machine tool drives and power factor con trol of synchronous generators

Regulex The Regulex (regulating exciter) is a dc generator acting as a power amplifier By proper design of the machine magnetic core an ex tensive linear portion of the voltage build up curve is obtained (Fig 10) A small change in mmf F will produce a large change in induced voltage E resulting in a degree of amplification Critical value adjustment of the field rheostat R (Fig 11) will cause the generator to operate on this linear portion A reference field F_2 and an opposing field F_3 combine with field F_1 to establish a point of op eration such as point a in Fig 10

Departure from this balance because of a varia tion in the control field F_3 will produce the large change in voltage E The output of this device may be used to drive a dc motor M which has its speed translated into voltage by means of a small pilot generator coupled to the motor shaft By proper feedback of this voltage to control field F_3 the mo- tor speed may be maintained at a constant value

Rototrol The Rototrol (rotating control) is a dc generator acting as a power amplifier It is similar to the Regulex but the self excited field is a series type in contrast to the shunt type field of the Regulex

Thymatrol The Thymatrol (thyratron motor control) is a complete electronic rectifier and con trol system in which a dc motor can be supplied with power from an ac source because of the recti fying action of thyratron tubes Several feedback circuits may be employed allowing automatic con trol of speed torque or current limitations

In Fig 13 the thyratron tubes furnish direct current to the motor armature in a full wave grid control rectifier circuit An increase in motor load will decrease the motor speed and the terminal voltage of the pilot generator, momentarily driving the grids of the thyratrons more positive This ac tion allows the thyratrons to fire at an earlier point in the positive cycle of the ac voltage applied to the plates increasing the average value of plate

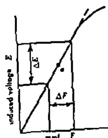


Fig 10 Regulex magnetization curve

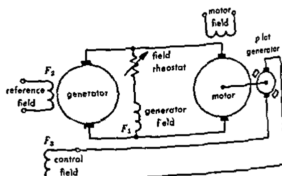


Fig 11 Typical Regulex circuit

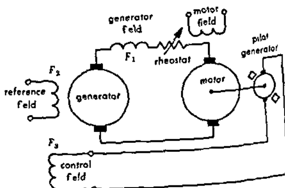


Fig 12 Typical Rototrol circuit

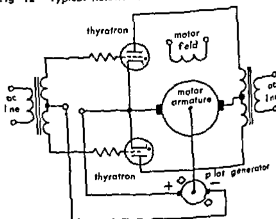


Fig 13 Elementary Thymatrol circuit

currents and the voltage across the motor armature, and restoring the speed A new point of voltage equilibrium in the feedback circuit must be maintained to sustain the original speed at the new load condition For discussion of thyratron tube operation, *see* THYRATRON

Mototrol The Mototrol (motor control) is similar to the Thymatrol [L F C]

Bibliography C L Dawes, *A Course in Electrical Engineering*, vol 1, 4th ed, 1952, A E Fitzgerald and C Kingsley, Jr, *Electric Machinery*, 1952, J G Truxal (ed), *Control Engineers' Handbook*, 1958

Directional coupler

A wave guide network of four guide terminals A, B C, and D (see Fig 1) such that there is a complete isolation between A and C and between B and D, but no isolation between the two terminals of any other combination The principle and usefulness of directional couplers will be illustrated by a simple example

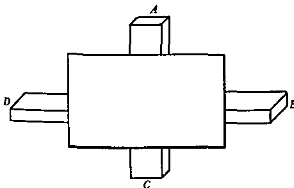


Fig 1 A network with four wave guide terminals

One common type of directional coupler is based on the interference of waves Consider two rectangular wave guides having their narrow sides joined as shown in Fig 2 Two small holes of the same size and shape are placed on their common narrow side spaced at a distance of one quarter wavelength ($\lambda/4$) apart Suppose a wave propagates into the network from terminal A Most of the wave will reach terminal B At either hole 1 or 2 a portion of the wave will leak into the second wave guide and divide into two equal but oppositely directed components Waves 1_R and 2_R will be combined in phase as they travel toward terminal D However, wave 2_L lags behind wave 1_L by 180° in phase, and no wave will reach terminal C because of cancellation Therefore there is a complete isolation between terminals A and C, as there is between terminals B and D by the same reasoning

In the example cited here, there is a large difference in the coupling factors between different pairs of terminals Thus terminals A and B (or C and D) are almost directly coupled with a coupling factor nearly equal to 1, whereas the coupling between A and D (or between B and C) can be made as small as desired A directional coupler of this type is usually rated by the coupling factor in decibels (typically 10 or 20)

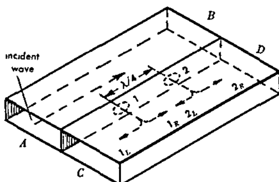


Fig 2 A two hole directional coupler

A magic tee junction is another form of directional coupler where the transmission coupling factor is ½ For a discussion of the properties of a magic tee and other wave guide junctions, *see* WAVE GUIDE

A directional coupler is most useful in selectively measuring either the forward (incident) or the backward (reflected) wave In the case of a small hole coupling (large decibel rating), this objective can be achieved with a negligible attrition of the main wave [C K J]

Direction-finding equipment

Radio aids to navigation that determine the direction of arrival of a radio signal by measuring the orientation of the wave front or of the magnetic or electric vector of a radio wave

Automatic direction finder. This device, commonly called ADF, is employed aboard aircraft to derive directional information from any constant (unkeyed) carrier ground radio station Radio stations erected specifically for use with the ADF are known as NDB (nondirectional beacons or radiophares) These have power outputs up to 3 kilowatts The ADF is commonly operated in the range of from 200 to 400 kilocycles (kc), but designs have been produced that operate from 90 to 2000 kc

The ADF presents to the pilot the direction of the radio station in relation to the aircraft Upon selecting a station, the needle of a bearing indicator automatically positions itself to indicate the direction of arrival of a radio wave

The common ADF was derived from the right left radio compass The principle of this device is illustrated by Fig 1 The radio compass employs both a loop antenna and a nondirectional antenna The loop antenna is mounted rigidly to the aircraft, with its plane at right angles to the longitudinal axis of the aircraft By proper phasing within the receiver, the receiving antenna pattern is a cardioid which can be reversed so that its maximum moves from the right to the left side of the aircraft as the connection to the loop antenna is reversed The output of the compass receiver

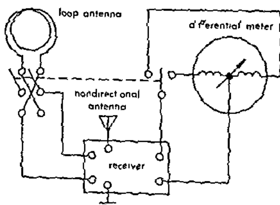


Fig 1 Principle of the right left radio compass

rectified and connected to a differential meter The meter reads the relative output for the two loop antenna connections Therefore it shows the relative strength of the signals received from one side of the aircraft as compared with those received from the other side If the radio station is directly ahead the reading of the meter will be zero This principle is illustrated in Fig 2 In actual practice the switching is accomplished by electronic methods at frequencies ranging from 30 to 90 cycles per second It may be considered that the operation consists of modulating the carrier frequencies as received by the antennas for the two loop connections with the switching frequencies The detector then contains two switching frequency components one at a phase advanced by an angle equal to the bearing of the station from the aircraft heading and the other at the same phase but minus an angle equal to the bearing of the station from the aircraft These components introduced into a meter having a field excited by the original modulation frequency produce a right or left deflection which is a function of the bearing angle

The ADF derived from the radio compass is illustrated by Fig 3 It may be considered that the

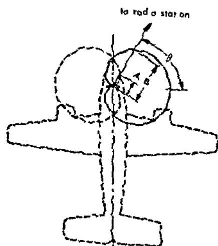


Fig 2 Use of compared cardioid patterns for indicating the heading of an aircraft with respect to a radio station

right left meter operates a reversible motor which positions the loop antenna so that the direction of the incoming signal is always at right angles to the plane of the loop Actually, the output of the receiver feeds a servomechanism which rotates the loop and incorporates feedback to produce zero damping

High frequency direction finder. This ground based radio direction finder, operating at frequencies of 2 to 20 megacycles is used mainly for

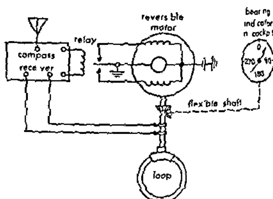


Fig 3 Principle of the automatic direction finder using a right left radio compass

navigational assistance in the long distance encounter zone

To obtain navigational information a pilot calls for a bearing by using his high frequency (hf) transmitter The ground station takes a bearing on the transmission and passes the information back to the aircraft via hf communications This navigational aid was very popular in Europe and is still employed for rescue purposes particularly with aircraft down at sea The hf ground direction finder has ... of hf propagation suc... deviation ... employ ... consists of four vertical elements spaced less than one half wavelength They are connected to a receiver by means of a goniometer The goniometer may be rotated manually or by a motor to obtain an indication on a cathode ray tube

One of these systems employs a twin path radio receiver The input of one receiver is connected to one pair of antennas and its output is connected to one pair of deflecting plates of a cathode ray tube The second receiver is similarly connected to the second antenna pair and the other set of deflecting plates Bearing accuracies of hf direction finders vary with propagation conditions Good bearing accuracy is considered to be about 5 degrees *See* NAVIGATION SYSTEMS ELECTRONIC

[ECS]

Bibliography D S Bond *Radio Direction Finders*, 1944, R Keen *Wireless Direction Finding* 3d ed 1938, P C Sandretto *Electronic Avigation Engineering* 1958

Directivity

The general property of directional discrimination displayed by systems which receive or emit waves Thus loudspeakers microphones radio antennas underwater hydrophones and even telescopes all have the common property that their effectiveness depends upon the direction from which the wave is either emitted or received The manner in which a sender or receiver is directional depends upon its geometrical shape and in particular upon its dimensions compared to the wavelength involved (designated here as λ)

Directivity is a desirable property of a receiver because it allows the identification of the direction from which a signal comes and because noise from other directions is eliminated It is desirable in a sender because the available energy may be concentrated in a given direction A simple example of directivity in a sender is a megaphone Audible sounds emitted from a person s mouth do not have significant directivity because the wavelengths involved are larger than the dimensions of the emitting area A megaphone effectively increases the size of the emitting area thereby increasing the directivity This raises the intensity of the sound in the forward direction at the expense of that in the backward direction The total emitted energy remains unchanged *See* MEGAPHONE

Directivity is specified in terms of plane waves of a given wavelength The directivity pattern of a receiver gives the relative response (either voltage or pressure) for plane harmonic waves arriving at different angles the direction of maximum response being assigned a value of unity For a sender directivity specifies the relative signal emitted as a function of angle the receiving distance being large compared to the size of the sender For a given frequency a system has the same directivity pattern used as either a sender or a receiver Thus in the following only the word receiver will be used to describe directivity

Directivity arises because various parts of the receiver respond to the incoming wave in different phase As an example consider a dipole receiver consisting of two elements each small compared to λ (and therefore omnidirectional by themselves) which are separated by a distance d A plane wave striking so that the wavefront is parallel to the axis joining the two is always in phase regardless of the value of λ Thus the response is always maximum for waves coming from directions normal to the axis The response for a wave coming along the axis depends on the ratio of d to λ If d $\lambda/2$ the response is zero because the two elements receive signals which are out of phase The polar directivity pattern for this case is shown in Fig 1*a* If more elements are added all spaced $\lambda/2$ apart directivity is increased and side lobes (or minor lobes) appear These are other directions of preferential response which are of lower response than the central direction This is illustrated in Fig 1*b*

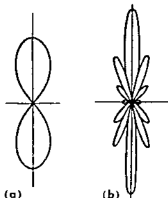


Fig 1 (a) Polar d rect v ty pattern for a d pole w th spac ng $\lambda/2$ (b) Polar d rect v ty pattern for l near array of seven elements spaced $\lambda/2$ apart

The cases just described are examples of linear arrays where directionality is obtained only with respect to one angular direction Directionality in both angles (that is where a beamlike pattern is formed) is obtained by using a receiver having extension in two dimensions A common example of this is a plane circular piston The directivity pattern for a piston is shown in Fig 2 for the case where the diameter is 5λ If the ratio of piston diameter to λ is increased the main response beam

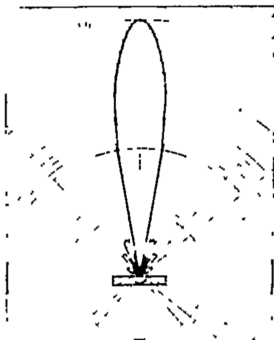


Fig 2 Polar d rect v ty pattern for plane c rcular piston w th d ameter of 5λ

is narrowed and the side lobes are reduced Patterns similar to this are found with sonar projectors and radar transmitters or receivers *See* ANTENNA (AERIAL) SOUND [R W MO]

Dirigible

A lighter than air craft equipped with means of propelling and steering for controlled flight in all three dimensions of the atmosphere As a general designation the term dirigible has gradually be come associated with the term airship and has

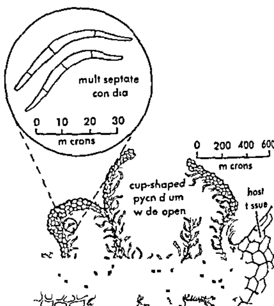
pelin type in which a complete framework of gird ers with diagonal wires and a doped fabric outer cover or metal skin provides structural strength without internal pressure The lifting gas is re tained in about 12 independent cells inside this structure arranged to give a compartment or bulk head system similar to that used by steamships The framework of the hull structure of a rigid airship consists of main and intermediate transverse frames usually of regular polygonal shape and in terconnected at their corners by longitudinal gird ers The rectangular or trapezoidal side panels formed by the girders of the transverse frames and the longitudinals on the surface of the hull are braced by a system of diagonal wires In the plane of the main frames are brace wires arranged radi ally or otherwise to form bulkheads which di vide the interior of the airship into compartments The intermediate frames have no transverse wiring They support the longitudinals and take stresses which result from gas pressure and outer cover loads The special function of the main frames is to carry the major part of the static loads such as car and engine weights usually located in the lower part of the airship to the aerostatic and aerody namic lifting loads that are distributed over the cir cumference When the airship noses up or down in flight or in case a cell becomes deflated the main frame bulkheads support the transverse area of the gas cells The longitudinal girders resist axial compression or tension caused by the main shear and bending of the airship as a body They are also subjected to local bending due to gas pres sure and outer cover loads The wiring system in the side panels gives the airship its shear strength and supports the gas cells between the girders *See* AEROSTAT [K AR R S R]

Discellaceae

A family of fungi also known as Excipulaceae of the order Sphaeropsidales Although most of the 225 species of the 70 genera reported are sapro phytes some are plant pathogens Many of the species are conidial stages of Discomycetes (sub class Ascomycetes)

The fruit body bearing conid a (pycnidium) is closed early in development but later opens out into a flack cup-shaped or disk shaped structure

Sporonema and *Dothichiza* are genera with 1 celled hyaline spores (spore group Hyalosporae) In *Sporonema* the pycnidia do not burst through the substratum *S phacidioides* causes leaf spot of alfalfa In *Dothichiza* the pycnidia break through



Pycn dia and conid a of *Brunchorst a destruens* (After C Ferd nandsen and C A Jorgensen 1938–1939)

the substratum and do not have openings *D (Chondroplea) populea* causes canker of poplar

Discella is a genus with 2 celled hyaline spores (Hyalodidymae) the disk shaped pycnidia are subepidermal *D carbonacea* in twigs of *Salix* is a stage of *Cryptodiaporthe salicina* (Ascomycetes)

Brunchorstia is a genus with hyaline spores which are 2 celled or more with cross septa (Hya lopragmiae) *Brunchorstia destruens* (*Rhabdospora pinea*) affects the shoots of pine trees

Heteropatella is a genus with hyaline 1 celled spirally coiled conidia that are long and curved at the tip (Hyaloscolecosporae) Some species are stages of *Heterosphaeria* (Ascomycetes) *H caltel linensis* affects carnations and *H antirrhini* ant r rhinums *See* FUNGI IMPERFECTI PLANT DISEASE SPHAEROPSIDALES [N F B]

Discriminant

For a polynomial $f(x) = a_n x^n + a_{n\ 1} x^{n\ 1} +$ $+$ $a_1 x + a_0$ the discriminant is given by the expres sion $D = a_n^{2n\ 1}(x_1 - x_2)^2 (x_1 - x_3)^2$ $(x_1 - x_n)^2$ $(x_2 - x_3)^2$ $(x_2 - x_n)^2$ $(x_{n\ 1} - x_n)^2$ where x_1 x_2 x_n are the roots of the equation $f(x) = 0$ There are $n(n-1)/2$ terms $(x_i - x_j)^2$ in the product corresponding to all possible selec tions of two indices i less than j from the numbers 1 2 n

The importance of the discriminant D lies in the fact that D vanishes if and only if the equation $f(x) = 0$ has equal roots (*see* EQUATIONS THEORY OF) Since the value of the discriminant is un changed if any two letters x_i and x_j are inter changed it is a symmetric function of the roots and can be expressed in terms of the coefficients of $f(x)$ Such an expression for D is most easily ob tained using the result $D = [(-1)^{n(n\ 1)\ 2}/a_n]$ $R_x(f\ f')$ where $f' = f'(x)$ is the derivative of $f(x)$

and $R_x(f,f')$ is the resultant of $f(x)$ and $f'(x)$ (*see* POLYNOMIAL SYSTEMS OF EQUATIONS) For example with $f(x) = a_2x^2 + a_1x + a_0$ and $f'(x) = 2a_2x + a_1$

$$R_x(f,f') = \begin{vmatrix} a_2 & a_1 & a_0 \\ 2a_2 & a_1 & 0 \\ 0 & 2a_2 & a_1 \end{vmatrix} = -a_2a_1^2 + 4a_2^2a_0$$

and $D = a_1^2 - 4a_2a_0$ which is the discriminant of a quadratic polynomial The discriminant of a cubic polynomial $f(x) = a_3x^3 + a_2x^2 + a_1x + a_0$ is

$$D = 18a_3a_2a_1a_0 - 4a_2^3a_0 + a_2^2a_1^2 - 4a_3a_1^3 - 27a_3^2a_0^2$$

[R A B]

Discriminator

A circuit that transforms a frequency modulated wave into a wave that is partially amplitude modulated and partially frequency modulated Appropriate filter circuits can be used to attenuate the residual frequency modulated wave leaving only an amplitude-modulated signal The remaining amplitude modulated signal is then detected in the usual manner by means of a linear or a square law detector The most common discriminator circuit is the phase-shift discriminator For this and other circuits *see* FREQUENCY MODULATION DETECTOR

[R L R]

Disease

A dynamic state in living organisms in which the normal characteristics of structure or function are altered The term disease is used in many ways and has been employed synonymously with illness sickness a *specific agent causing illness and so on*

The healthy body or tissue tends to maintain homeostasis that is a steady state of acceptable physiologic activity by means of many regulatory mechanisms When any extrinsic or intrinsic factor causes an alteration of the body or tissue so that the available homeostatic mechanisms cannot overcome the alteration disease results

In another sense disease is the failure of the body or one of its parts to adapt to a change Such failure in adaptation may be marked by changes in the morphology or function of the affected part in either a manner characteristic of the process or a nonspecific way In certain cases failures of adaptation may produce no detectable local changes although the total effect of the alteration is obvious

The word disease may also be used to indicate a rather specific course of events that can be attributed to the alterations produced by certain agents that affect the body or its parts Thus the diseases of rheumatic fever or arteriosclerosis represent recognizable changes or failures of adaptation that are characteristic of these processes The alterations of tissues in such diseases are known as the lesions or the pathologic lesions They represent the reactions of cells tissues and organs to injury The factor or factors which bring about such alterations are called the etiologic agents In many cases they are well known as in tuberculosis bacillary dysentery and poliomyelitis In other instances the etiologic factors may be multiple uncertain or unknown *See* BACILLARY DYSENTERY POLIOMYELITIS RHEUMATIC FEVER TUBERCULOSIS

The terms sickness and illness are used most often by laymen to indicate the more subjective nature of some body disorder In the medical sense the term disease indicates an attempt to be more objective in the consideration of etiology incidence clinical course pathologic lesions and related components of a failure of adaptation

It should be recognized that there is no sharp line between extremes of physiologic response to stimuli and the often ill defined beginnings of a disease process Consideration must be given to the individual—his sex age habits health pattern genetic background and many other pertinent factors

[E C ST]

BIOCHEMISTRY

The chemical composition of the body includes proteins lipids carbohydrates inorganic salts water enzymes vitamins and hormones Disease presupposes a shifting of the normal balance of these substances and usually a decrease or an increase in their quantity

Diseases due to hormones Diabetes is a disturbance in carbohydrate metabolism In part at least it is due to an insufficient supply of the hormone insulin (in the pancreas) Insulin is a protein whose chemical contribution has been elucidated by F Sanger Besides insulin glycogen another hormone in the pancreas and a hormone of the anterior pituitary are involved in the disease *See* HORMONE

The thyroid made up of two lobes one on each side of the trachea and the larynx contains among others the hormone thyroxin (an iodinated phenol derivative) which regulates the rate of metabolism Cretinism in the young and myxedema among older people are examples of thyroid insufficiency and exophthalmic goiter is an example of hyperthyroidism *See* THYROID GLAND

The parathyroids four small organs attached to the thyroid develop a hormone which regulates the metabolism of calcium Tetany develops as the blood calcium falls from its normal value *See* PARATHYROID GLAND

The pituitary is situated at the base of the skull and has a multiplicity of functions The anterior pituitary is responsible for the following hormones which in many cases have been isolated and have been shown to be protein in nature growth gonadotropic (interstitial and follicle stimulating) lactogenic thyrotropic and adrenotropic Increased or decreased amounts of these hormones cause serious disturbances throughout the body

Two substances in the posterior portion of the pituitary gland have been isolated vasopressin an antidiuretic and oxytocin which contracts the muscles of the uterus Both of these substances peptides have been isolated and synthesized by V Du Vigneaud

The adrenals (situated near the kidneys) have two parts the medulla and the cortex The medulla contains the hormones epinephrine (adrenalin) and norepinephrine derivatives of catechol The cortex contains several hormones which have been shown to be related to the sterols, deficiency of some of them gives rise to Addison s disease

The sex hormones in the testes and the ovaries are secreted under the stimulus of hormones from the anterior pituitary The genital tract and the accessory male organs are influenced by the male hormones (for example testosterone) In the female one type of hormone estradiol is a product of the ovary another progesterone is derived from the corpus luteum All the natural sex hormones are sterol compounds

The hormones are occasionally useful in various sex dysfunctions

Some researchers believe that hormones play a part in carcinogenesis

Vitamin deficiency diseases A vitamin is an organic compound—other than protein fat or carbohydrate—which belongs to the group of essential foodstuffs Many of these vitamins have been isolated and synthesized Their chemical structure covers a wide range The diseases due to lack of vitamin (avitaminoses) are shown in the table *See* VITAMIN

Avitaminoses

Vitamin	Vitamin deficiency effects
Vitamin A	Loss of weight decreased resistance to infection xerophthalmia (eyes incrusted and infected)
Thiamine	Interference with glycogen storage
Riboflavin	Cheilosis (reddening of the eyes) glossitis (inflammation of the tongue)
p-Aminobenzoic acid	?
Niacin	Pellagra
Vitamin B_6 (pyridoxine)	Acrodynia (dermatitis)
Pantothenic acid	Dermatitis graying hair
Biotin	Spectacle eye alopecia (baldness)
Inositol	Alopecia
Choline	Fatty liver
Folic acid analogs	Anemia
Vitamin B_{12}	Pernicious anemia
Ascorbic acid	Scurvy
Vitamin D	Rickets
Vitamin E (tocopherols)	Sterility?
Vitamin K	Hemorrhagic disease

Inborn errors of metabolism There are several well recognized diseases which are hereditary and involve the absence of a specific enzyme The enzyme is normally required in a sequence of reactions The absence of one of these enzymes results in the disease alkaptonuria in which there are large quantities of homogentisic acid in the urine Tyrosinosis is a disease in which *p* hydroxyphenyl pyruvic acid is excreted Phenylketonuria is a disease in which phenylpyruvic acid appears in unusual quantities in the urine Albinism is a disease in which dopa (3 4-dihydoxyphenylalanine) fails to be oxidized to the melanin pigments *See* HUMAN GENETICS

Miscellaneous. Lipoproteins—complex compounds of lipids and proteins—may be involved in the disease known as atherosclerosis

It has been maintained that fluorine is related to resistance to dental caries Drinking water containing one part per million of fluorine (as sodium or potassium fluoride) is advocated

Gout a disease of unknown origin reveals a slow accumulation of uric acid (as sodium biurate)

[B H]

Bibliography R L Cecil and R F Loeb *A Textbook of Medicine*, 1955, B Harrow and A Mazur *Textbook of Biochemistry*, 1958

Disease carrier

A person who is infected with certain microorganisms and is therefore liable to transmit the disease produced by them but who does not himself show signs or symptoms of the infection He may be in such an early phase of the disease that clinical manifestations are not apparent and in this case he is an incubatory carrier If he has already had a clinical attack, he may be described as a convalescent carrier These carriers sometimes continue to harbor the infectious agent for such an extended period that they are called chronic carriers It is also possible for a person to carry microorganisms even though he does not at any time manifest the disease, this association with the germ is an incidental one and the person is described as a casual carrier *See* DISEASE

The existence of healthy carriers is a complicating factor in the study of infectious diseases Since they are without symptoms or signs of disease they can be identified as carriers only by laboratory investigation In poliomyelitis for every clinical case there are probably several hundred who are carriers of the virus but do not themselves suffer from it

Patients who have had typhoid fever are particularly liable to become chronic carriers harboring the organism in the gallbladder and excreting it in the feces *See* EPIDEMIOLOGY, TYPHOID FEVER

[C W H]

Disk recording

The process of inscribing suitably transformed acoustical or electrical signals on a flat circular plate that may be played back at a subsequent time Virtually all modern disk recorders and reproducers are used to record or reproduce sound signals mainly music and voice

This article discusses both monophonic and stereophonic recording and reproducing systems and their compatibility commercial phonographs the manufacture and specifications of disk records and distortion and noise in record reproduction For related information *see* AMPLIFIER HIGH FIDELITY LOUDSPEAKER, MICROPHONE, SOUND REPRODUCTION SYSTEMS ELECTRICAL, *see also* MAGNETIC RECORDING OPTICAL RECORDING

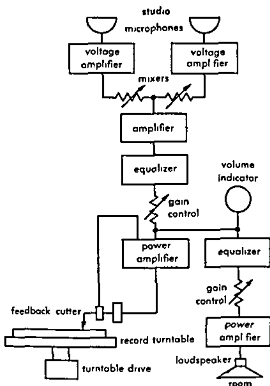


Fig 1 Schematic arrangement of apparatus in a complete monophonic disk recording system

MONOPHONIC SYSTEM

A monophonic disk recording system consists of a disk record rotated by a turntable mechanism and a cutter for producing undulations in a groove in the disk corresponding to the sound signals A monophonic disk reproducing system consists of a pickup and mechanism for rotating the disk record by means of which the recorded undulations in the disk record are converted into electrical signals of approximately like form

Recording system. The elements in a complete monophonic disk recording system are shown in Fig 1 The first element is the acoustics of the studio The output of each microphone is amplified and fed to a mixer, a device having two or more inputs and a common output If more than one microphone is used (for example when an orchestra accompanies a soloist, there is one microphone for the orchestra and one for the singer), the outputs of the two microphones may be adjusted for the proper balance by means of the mixers An electronic compressor is used to reduce a large amplitude range to that suitable for reproduction in the home A corrective electrical network called an equalizer provides the recording characteristics which are described later The attenuator, or gain control, provides a given control on the over all level fed to the power amplifier The cutter, actuated by the amplifier, cuts a wavy path in the groove of the revolving record corresponding to the undulations in the original sound wave striking the microphone A monitoring system consisting of a volume indicator, complementary equalizer, attenuator or gain control power amplifier, and loud speaker is used to control the recording operation The volume indicator employs a logarithmic or decibel scale calibrated in volume units (VU) The volume unit is defined as 10 times the common logarithm of the power ratio p_2/p_1, where the reference power level p_1 is selected as 1 mw (0 001 watt) and p_2 is the signal power level

Recorder A phonograph recorder is an instrument for transforming acoustical or electrical signals into motion of approximately like form and inscribing such motion in an appropriate medium by cutting or embossing For the recording of disk phonograph records the electrical phonograph recorder (Fig 2) replaced the mechanical recorder in the late 1920s The lacquer disk used in recording the master record is placed on the recording turntable The turntable is heavy, to ensure against spurious rotational motions A suitable mechanical filter is placed between the driving motor and the turntable so that uniform rotational motion of the turntable will be obtained The drive system is arranged so that records of all standard speeds can be cut In general the recording turntable is driven with a synchronous motor to ensure constant speed of rotation The lead screw drives the cutter in a radial direction so that a spiral groove is cut in the record Lead screws of different pitches are used ranging from 100 to 500 grooves per inch In some recordings a variable pitch is used In this procedure the spacing between the grooves is made to correspond to the amplitude—small spacing for small amplitudes and large spacing for large amplitudes Under these conditions the maximum amount of information can be recorded on a record The material removed in the cutting process in the form of a fine thread is pulled into the open end of a pipe which is located near the cutting stylus and is connected to a vacuum system

Cutters The electromechanical transducers used as cutters can be of either the lateral or the vertical type In the lateral type of disk recording, the undulations are cut in a direction parallel to the

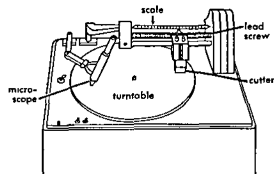


Fig 2 Perspective view of disk phonograph recorder The microscope is used for periodic inspection of the groove (*After H F Olson, Acoustical Engineering, Van Nostrand 1957*)

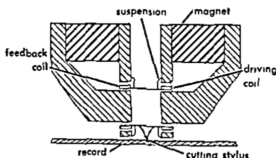


Fig 3 Sectional view of a lateral feedback phonograph cutter

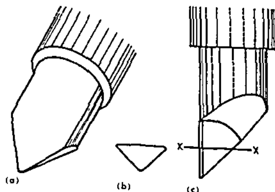


Fig 4 (a) Perspective (b) section and (c) side views of a cutting stylus

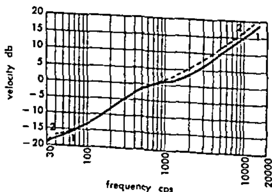


Fig 5 Velocity frequency recording characteristics used in commercial records Curve 1 Record Industry Association of America standard Curve 2 Orthacoustic standard (*From L C Smeby, Recording and reproducing standards Proc IRE 30(8) 355–356 1942*)

surface of the record and perpendicular to the groove A sectional view of a lateral feedback phonograph [illegible] system [illegible] the dri[illegible] [illegible]se [illegible]ing (feed back) coil wound on a common cylinder The cutting stylus is attached to the coil cylinder The vibrating system is designed so that it exhibits a single degree of freedom (single type of movement) over the frequency range from 30 to 16 000 cycles with a fundamental resonant frequency at 700 cycles The output of the sensing coil is fed to the input of the amplifier, as shown in Fig 1 The output of the amplifier is fed to the driving coil in an out of phase relationship With the feedback in operation, the velocity of the driving system is practically independent of frequency over the range from 30 to 16,000 cycles The input to the amplifier is compensated to provide the desired recording characteristic

The cutting stylus consists of a sapphire synthetic ruby, or other hard material fashioned in the form of a pointed chisel (Fig 4) The stylus is heated in recording and thereby imparts a smooth sidewall to the groove This expedient results in considerable reduction in noise in reproduction The stylus may be heated by a few turns of fine wire wound around it and operated from a low voltage dc supply

The original recording of disk records is made on a lacquer disk The lacquer disk consists of a coating of an acetate plastic on two sides of an aluminum disk The grooves are cut in the plastic

In the vertical (hill and dale) type of disk recording the undulations of the groove are cut in a direction perpendicular to the surface of the record and perpendicular to the groove The vertical cutter is similar to the lateral cutter except that the stylus is located on the end of the cylinder of Fig 3 and the entire system is turned 90°

The vertical disk phonograph system is used to a limited extent in broadcasting stations but it is not used in home disk phonograph systems

Recording characteristics The velocity-frequency response characteristic of the groove in the phonograph disk record provides the velocity at the tip of the stylus of the phonograph pickup as a function of the frequency

Electrical transcriptions (a term applied to professional disk recordings) are cut with an orthacoustic recording characteristic on records 16 in in diameter turning at 33⅓ rpm The orthacoustic velocity frequency characteristic for constant voltage input to the microphone voltage amplifier is shown in Fig 5 This characteristic is essentially a constant amplitude frequency characteristic In reproduction of the disk record an inverse frequency response characteristic is used to obtain a uniform over all frequency response characteristic

In the recording of commercial phonograph records some form of high frequency compensation has always been employed The compensation that has been used since the advent of the disk phonograph has varied over wide limits Fortunately in 1954 the Record Industry Association of America standardized the velocity–frequency response characteristic of the groove in the commercial lateral disk record The RIAA standard velocity frequency response characteristic is shown in Fig 5 In the reproduction of commercial disk phonograph records an inverse frequency response characteristic is employed in order to obtain a uniform over all frequency response characteristic Commercial standard frequency records exhibiting the RIAA frequency response characteristic are used

in the development, design, and service of disk phonograph instruments. See EQUALIZATION; FREQUENCY RESPONSE.

Record manufacture. The processes for the mass production of disk phonograph records are depicted in Fig. 6. The original lacquer disk, termed the original, is metallized and then electroplated. The plating is separated from the lacquer and reinforced by backing with a solid metal plate. The assembly, called the master, is electroplated. This plating is separated from the master and reinforced by backing with a solid plate. The resulting assembly, the mother, is electroplated and reinforced by a solid metal plate, forming an assembly termed the stamper. Several stampers are made from each mother. One stamper containing a sound selection to be placed on one side of the final record is mounted in the upper jaw, and another stamper containing a sound selection to be placed on the other side of the record is placed in the lower jaw of a hydraulic press equipped with means for heating and cooling the stampers. A preform, or biscuit, of thermoplastic material such as a shellac compound or Vinylite is placed between the two stampers.

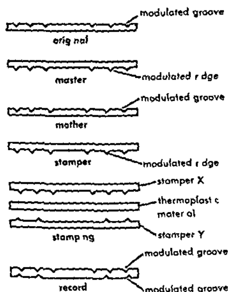


Fig. 6. Steps in mass production of disk phonograph records from lacquer originals.

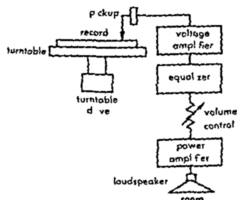


Fig. 7. Schematic arrangement of apparatus in a complete monophonic disk sound reproducing system.

The stampers are heated and the jaws of the press are closed to bring the two stampers against the thermoplastic material. When an impression of the stampers has been obtained in the thermoplastic material, the stampers are cooled, thus cooling and setting the plastic record. The jaws of the hydraulic press are opened and the record is removed from the press. The modulated grooves in the record correspond to those in the original lacquer disk. The stamping procedure is repeated again and again until the desired number of records is obtained. This process constitutes the mass production technique for the production of phonograph records.

Reproducing system. The elements of a complete monophonic disk sound reproducing system are shown in Fig. 7. The first element is the motor-driven turntable, which turns the record at a constant rotational speed. The stylus or needle of the pickup follows the wavy groove in the record, and the pickup transducer generates a voltage corresponding to the undulations in the record. The output of the pickup is amplified by a voltage amplifier. The amplifier is followed by an equalizer which complements the recording characteristic of the record. Filters and tone controls are provided for further equalization of the response according to the taste of the listener. The tone controls provide means for increasing or decreasing the low- and high-frequency response. In general, the increase or decrease in response starts at 1000 cycles with a gradual increase or decrease in both high and low directions. The maximum increase or decrease in response at the extreme ends of the frequency range covered is about ±15 decibels. A volume control provides means for obtaining the desired level of sound in reproduction. The volume control is followed by an amplifier which drives the loudspeaker.

Turntable and record changer. An electrical record player and changer is shown in Fig. 8. The record is rotated by the turntable at the same angular speed as that used in recording. The turntable is rotated by means of an electric motor. The stylus or needle of the pickup follows the wavy spiral groove and generates a voltage corresponding to the undulations in the groove. (Pickups used in disk record reproduction are described later.) The record player shown in Fig. 8 will play recordings at four rotational speeds, $16\frac{2}{3}$, $33\frac{1}{3}$, 45, and 78 rpm. It will also play and change a stack of eight records. The small spindle is used for $16\frac{2}{3}$, $33\frac{1}{3}$, and 78 rpm records. The large spindle is used for $16\frac{2}{3}$ and 45 rpm records.

Another type of record changer and player plays and changes a single type of record. One of the most common is the 45 rpm record player and changer.

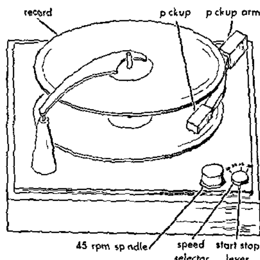


Fig 8 Perspective view of a four speed d sk phono graph player (*After H F Olson Acoustical Engineer ing Van Nostrand 1957*)

A record player is the simplest type of disk record reproducer It is manually operated It ranges from the simplest of all disk record players to elaborate transcription types with very uniform rotational velocity and high quality pickups

Pickups A phonograph pickup is an electromechanical transducer actuated by a phonograph record and delivering energy to an electrical system the electric current having frequency components corresponding to those of the wave in the record The following systems are used for converting the mechanical vibrations into the corresponding electrical variations magnetic condenser electronic dynamic ceramic and crystal

Modern lateral pickups are of the crystal ceramic magnetic or dynamic type A crystal phonograph pickup depends for its operation on the piezoelectric effect (see PIEZOELECTRICITY) The crystal used is Rochelle salt A cross sectional view of a typical crystal pickup used for commercial phonographs is shown in Fig 9 The stylus driven by the record is coupled through an arm to the crystal and thereby produces a twist in the crystal The open circuit output of the crystal is proportional to the twist or displacement The open circuit voltage of a crystal pickup for an amplitude of 0 001 in is about 0 5 volt The open circuit voltage displacement characteristic makes the frequency equalization exceedingly simple because the recording characteristic shown in Fig 5 exhibits a practically constant displacement frequency characteristic The electrical capacitance of the crystal is of the order of 1000–2000 micro microfarads The equivalent electrical circuit is the open circuit voltage in series with the electrical capacitance The electrical impedance presented to the crystal pickup must be larger than the electrical impedance of the crystal in orde

... prevent frequency discrimination against the low frequency range

A ceramic phonograph pickup depends for its operation on the electrostrictive effect (see ELECTROSTRICTION) The ceramic used is barium titanate The ceramic pickup may be made in designs similar to that of the crystal pickup The characteristics are similar in all essential respects save that the sensitivity is somewhat lower than that of the Rochelle salt crystal

A phonograph pickup which depends for its operation on the variation in magnetic flux through a stationary coil is called a magnetic pickup (see MAGNETIC FLUX) A modern type is shown in Fig 10 The horizontal stylus also serves as the armature This design makes it possible to obtain a relatively low mechanical impedance The steady flux is supplied by a small permanent magnet As the armature is deflected from the central position the flux through one coil is increased and the flux through the other coil is decreased The coils are connected in series so that the resultant voltages generated in the two coils are added The open cir

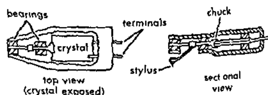


Fig 9 Top and sectional views of crystal phonograph pickup

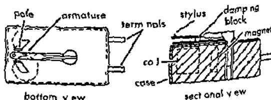


Fig 10 Bottom and sectional v ews of magnetic phonograph p ckup

cuit voltage generated in a coil is proportional to the time rate of change of magnetic flux through the coil which in turn is proportional to the velocity of the armature Thus the open circuit voltage generated in the coil will be independent of the frequency if the velocity of the armature is independent of the frequency In a properly designed magnetic pickup the open circuit voltage frequency response characteristic will correspond to the groove velocity frequency response characteristic For the recording characteristic of Fig 5 frequency compensating networks must be employed in order to obtain a uniform output frequency response characteristic The output of a typical magnetic pickup for an amplitude of 0 001 in at 1000 cycles is of the order of 0 010 volt The electrical impedance is highly inductive and therefore the electrical impedance is nearly proportional to the frequency The electrical impedance of a

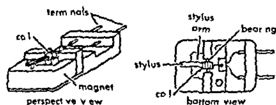


Fig 11 Perspective and bottom views of dynamic pickup

typical magnetic pickup for a 0 010-volt output is 2500 ohms at 1000 cycles The equivalent electrical circuit is the open circuit voltage in series with the electrical impedance of the coils

A dynamic phonograph pickup depends for its operation on the motion of a conductor in a magnetic field (see INDUCTION ELECTROMAGNETIC) A typical dynamic pickup is shown in Fig 11 The stylus arm is coupled to a coil located in a magnetic field The open circuit voltage developed in the coil is proportional to the rate of change of magnetic flux through the coil and will be independent of frequency if the velocity is independent of frequency In a properly designed dynamic pickup the open circuit voltage frequency response characteristic will correspond to the groove velocity frequency response characteristic For the recording characteristic of Fig 5 frequency compensating networks must be employed in order to obtain a uniform output frequency response characteristic Since the vibrating system may be very small and light it is possible to obtain a uniform velocity frequency response characteristic over a wide frequency range The electrical impedance is practically an electrical resistance of about 25 ohms The output at 1000 cycles for an amplitude of 0 001 in is about 0 001 volt In general a transformer is used to step up the output voltage and electrical impedance

In the vertical type of disk recording the undulations of the groove are perpendicular both to the plane of the disk and to the groove Therefore the vibration of the stylus of the pickup is in the vertical direction Any of the lateral pickups just described may be used by turning the transducer 90°

Groove and stylus dimensions A sectional view of the groove of a lateral disk record and the stylus is shown in Fig 12 The width of the groove at the surface of the record A the angle of the walls of the groove G the radius C of the stylus the distance B below the surface of the record to the contact point of the stylus with the groove and the width of the contact points are shown in the figure The dimensions and angles for a coarse groove fine groove and ultrafine groove are given in the table The coarse groove is used in 78-rpm records the fine groove in 45- and 33⅓-rpm records and the ultrafine groove in 16⅔ rpm records

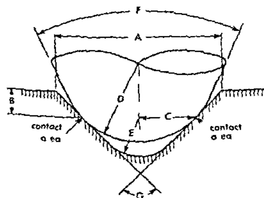


Fig 12 Sectional view of stylus in a groove (After H F Olson Acoustical Engineering Van Nostrand 1957)

Values of dimensions and angles of Fig. 12

	Coarse groove	Fine groove	Ultrafine groove
A in	0 006	0 00[illegible]	0 001
B in	0 0008	0 0004	0 0004
C, in	0 0019	0 0007	0 00017
D in	0 00[illegible]7	0 001	0 000[illegible]5
E, in	0 0023	0 000[illegible]7	0 00015
F	45°	4[illegible]°	4[illegible]°
G	90°	90°	90°

The maximum nominal grooves per inch for the different sized grooves are as follows coarse groove 125 fine groove 275 and ultrafine groove 550

The maximum amplitudes in inches in the frequency range 200–2000 cycles for the different sized grooves are as follows coarse groove 0 004–0 005 in fine groove, 0 0015–0 002 in and ultra fine groove 0 0007–0 001 in

COMMERCIAL DISK RECORDS

Commercial phonograph records are made in four speeds namely 78 (approximately) 45 33⅓, and 16⅔ rpm The 78-rpm records are made in three diameters 12, 10 and 7 in The normal maximum playing times for full width records are 5 3½ and 2¼ min, respectively The 33⅓ rpm records are made in three diameters 12 10 and 7 in The normal maximum playing times are 25 17 and 8 min respectively The 45-rpm records are made in a diameter of 7 in and have a normal maximum playing time of 8 min The 16⅔ rpm records are made in a diameter of 7 in The normal maximum playing time of the records with the large center hole is 30 min while that for the small hole records is 45 min for music and 60 min for speech The over all diameter the diameters of the

preceding discussion are representative and do not include all the variations

DISTORTION AND NOISE

Distortion in reproduction The recording and reproducing of a phonograph record constitute a complicated process, with many sources of non ear distortion The record does not present an

infinite mechanical impedance to the stylus. As a consequence, the vibrating system of the pickup is shunted by the effective mechanical impedance of the record at the stylus. Nonlinear distortion will be introduced if this impedance of the record is variable. Other sources of distortion are tracking error and tracing distortion.

Tracking distortion. A nonlinear distortion due to a deviation in tracking is commonly termed tracking error. The angle between the vertical plane containing the vibration axis of the pickup and the vertical plane containing the tangent to the record is a measure of the tracking error. If the vibration axis of the pickup passes through the tone arm pivot, the tracking error can be zero for only one point on the record. Tracking error can be reduced if the vibration axis of the pickup is set at an appropriate angle with respect to the line connecting the stylus point and the tone arm pivot, together with provisions for a suitable overhang distance between the stylus and the record axis.

Tracing distortion. A form of distortion in lateral disk record reproduction known as tracing distortion is a function of the diameter of the stylus, the lateral velocity, and the linear groove velocity. This distortion is due to the fact that there is not a one-to-one correspondence between the shapes of the cutting and reproducing styli (Figs. 4 and 12). The cutting stylus presents a triangular shape as it cuts the groove. Thus it is seen that the groove narrows as the cutting stylus approaches the center position because the cutting stylus is moving at an angle with respect to the motion of the record. The reproducing stylus presents a spherical surface to the groove; therefore, as the reproducing stylus moves in the groove, it will rise as the groove narrows. The frequency of the rise is twice the fre

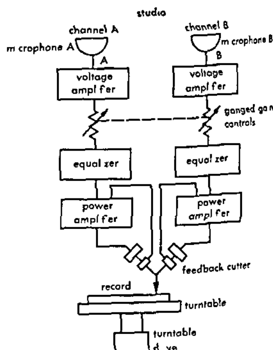


Fig. 14. Schematic arrangement of apparatus in a complete stereophonic disk recording system.

Fig. 13. Typical dimensions of the most common commercial disk phonograph records. (After H. F. Olson, *Acoustical Engineering*, Van Nostrand, 1957.)

quency of the modulation The narrowing of the groove is termed the pinch effect The two sides of the groove are symmetrical therefore the stylus must execute symmetrical motion about the center line which means that there should be no even harmonics However odd harmonics are produced

Other distortions If the force which the stylus presents to the record is of such magnitude that it exceeds the yield point of the record material the mechanical impedance of the record will not be a constant The result is production of nonlinear distortion Furthermore if the force exceeds the yield point by a considerable amount the record may be permanently damaged

As the needle or stylus is worn by the groove the shape of the point changes from a spherical surface to a wedge shape The wedge shaped stylus introduces nonlinear distortion and a loss in the high frequency response

A consideration of the load and needle forces at the stylus tip shows that there is a force which is proportional to the tracking angle This force known as side thrust is usually directed toward the center of the record and is applied to the inner boundary of the record groove it is responsible for the unequal wear on the two sides of the stylus

Another source of distortion results from the lack of correspondence between the linear groove speed in the recording and the ultimate reproduction This type of distortion which leads to a frequency modulation of the reproduced signal is termed *wow* Wow may be due to a nonuniform speed of the record turntable during recording or reproduction misplacement of the center hole or configuration distortion during the processing In general the major source of wow is the nonuniform speed of the reproducing turntable *See* FLUTTER AND WOW

Surface noise The record surface noise in the absence of any signal is one of the factors which limits the volume range and the frequency range of shellac phonograph records The amount of surface noise for a given record is proportional to the frequency bandwidth In order to reduce the surface noise to a tolerable value in shellac records it is usually necessary to limit the high frequency range in reproduction A method of decreasing the effective surface noise consists of increasing the amplitude of the high frequency response in recording and introducing complementary equalization as shown in the recording characteristic of Fig 5 The noise of Vinylite records is extremely low and in general is not a problem Adequate volume ranges can be obtained with Vinylite and other similar plastics

STEREOPHONIC SYSTEM

Two-channel disk phonograph sound reproduction was commercialized in 1958 The stereophonic disk phonograph provides the reproduction of the original sound sources in auditory perspective that is the spatial relations of the original sound are substantially retained in the reproduction of the recorded sound

Recording system The elements of a complete stereophonic disk recording system are shown in Fig 14 There are two channels identical to the type shown in Fig 1 except for the cutter A two channel disk phonograph dynamic type feedback cutter is shown in Fig 15 The two vibrating systems are arranged at right angles therefore the two channels in the groove are recorded at right angles The modes of vibration in a plane normal to the surface of the record and normal to the groove axis are shown in Fig 16 The motion of the cutting stylus is also shown in Fig 16

The same type of recorder described in the beginning of this article may be used in the recording of stereophonic records

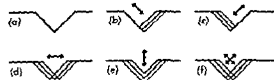


Fig 16 Schematic views of groove undulations in stereophonic disk phonograph system Heavy line indicates zero amplitude or unmodulated groove Light lines indicate maximum limit of groove modulation Arrows indicate direction of motion of recording cutter and reproducing stylus (a) Unmodulated groove (b) Modulation in right channel (c) Modulation in left channel (d) Lateral modulation combination of (b) and (c) in phase (e) Vertical modulation combination of (b) and (c) out of phase (f) Combination of equal vertical and lateral amplitudes combination of (b) and (c) with 90° phase shift

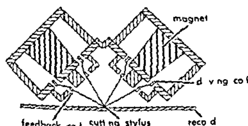


Fig 15 Sectional view of a feedback stereophonic disk phonograph cutter

The recording characteristics employed in stereophonic disk recording are essentially the same as those used in monophonic disk recording

Reproducing system The elements of a complete stereophonic disk reproducing system are shown in Fig 17 There are two identical channels following the pickup of the type shown in Fig 7 A two-channel disk phonograph dynamic pickup is shown in Fig 18 Each element of the stereophonic pickup consists of a transducer of the type employed in the single channel lateral dynamic pickup shown in Fig 11 Reference to Figs 16 and 18 shows that a vibration which excites one element not excite the other Other types of picku

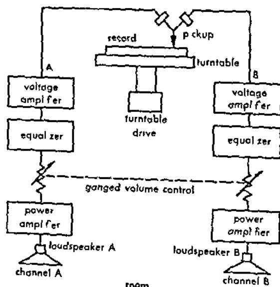


Fig 17 Schematic arrangement of apparatus in a complete stereophonic disk phonograph reproducing system

have also been developed For example in a ceramic pickup two ceramic elements are arranged with the vibrating planes at right angles and coupled to the stylus in such a manner that a vibration which excites one element will not excite the other

The groove used for the stereophonic disk record is the fine groove depicted in Fig 12 A stylus with a tip radius of 0 00075 in is recommended for use in reproducing stereophonic disk records

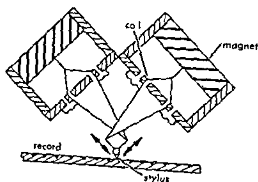


Fig 18 Sectional view of stereophonic disk phonograph pickup

Compatibility Monophonic and stereophonic records and phonographs are compatible as follows

When a single channel monophonic record is reproduced by a single channel monophonic phonograph system the single channel output is reproduced by the loudspeaker

When a single channel monophonic record is reproduced by the two-channel stereophonic phonograph system the single channel output is reproduced by the two loudspeakers

When a two channel stereophonic record is reproduced by a single channel monophonic phonograph reproducing system the sound reproduced by the single loudspeaker is the sum of the two sound programs originally recorded on the two channels of the stereophonic recording system

When the two channel stereophonic record is reproduced by a two channel stereophonic reproducing system the stereophonic sound program is reproduced on the separate channels corresponding to the recording channels and the sound which emanates from the two loudspeakers corresponds to the sound picked up by the respective microphones [H F O]

Bibliography J G Frayne and H Wolfe *Elements of Sound Recording* 1949 A Jorysz Bibliography of disk recording 1921–1947 *J Audio Engr Soc* 2(2) 92–108 1954, H F Olson, *Acoustical Engineering* 1957

Dispersion (radiation)

The separation of a complex electromagnetic or sound wave into its various frequency components For example a beam of white light can be separated into its monochromatic components by virtue of the different velocities of rays of different wavelength of the beam as it passes through a prism or grating The dispersion of a material such as glass or water at a given wavelength in the electromagnetic spectrum is defined as the rate of change of refractive index with wavelength at the wavelength in question For an extended discussion of the dispersion of light *see* ABSORPTION (ELECTROMAGNETIC RADIATION) [W W]

Displacement (mechanics)

When an object is moved from one position to another it is said to be displaced and the linear distance it is moved in a given direction is called the displacement The displacement is always in a particular direction consequently displacement is a vector quantity involving direction as well as magnitude The magnitude of any linear displacement is called the length of path or the distance traversed and since length of path does not involve direction it is a scalar quantity For instance the distance or length of path is the same from Washington to Philadelphia as it is from Philadelphia to Washington but the displacement of an object in one direction is entirely different from its displacement in the opposite direction

When a body is rotated about any axis it is said to undergo angular displacement Angular displacement is commonly measured in radians or degrees one radian being equal to 57.3° *See* ROTATIONAL MOTION [R D RI]

Displacement current

The name given by J C Maxwell to the term $\partial D/\partial t$ which must be added to the current density i to extend to time varying fields A M Ampere's mag

netostatic result that $\mathbf{i}$ equals the curl of the magnetic intensity $\mathbf{H}$ In integral form this result is

$$\oint \mathbf{H}\cdot d\mathbf{s} = \int_S \left(\mathbf{i}+\frac{\partial \mathbf{D}}{\partial t}\right)\cdot \mathbf{n}\, dS \qquad (1)$$

where the unit vector $\mathbf{n}$ is perpendicular to the surface dS The concept of displacement current has important consequences for insulators and for free space where $\mathbf{i}$ vanishes For conductors however the difference between Eq (1) and Ampere s result is negligible *See* AMPERE S LAW MAXWELL S EQUATIONS

In order to show that the displacement term is essential consider a parallel plate capacitor charged by a circuit carrying an alternating current Let a closed curve s encircle one of the charging wires and be the boundary of two surfaces S_1 which passes through the capacitor gap and S_2 which cuts the charging wire By Gauss electric flux theorem the charge Q on the plate and wire between S_1 and S_2 is

$$Q = \int_{S_1} \mathbf{D}\cdot \mathbf{n}\, dS - \int_{S_2} \mathbf{D}\cdot \mathbf{n}\, dS \qquad (2)$$

where the normal is taken in the direction of current flow in both S_1 and S_2 The current I or $\int \mathbf{i}\cdot\mathbf{n}\, dS$ equals $\partial Q/\partial t$ so that

$$\int_{S_1} \frac{\partial \mathbf{D}}{\partial t}\cdot \mathbf{n}\, dS = \int_{S_2} \left(\mathbf{i}+\frac{\partial \mathbf{D}}{\partial t}\right)\cdot \mathbf{n}\, dS = \oint \mathbf{H}\cdot d\mathbf{s} \qquad (3)$$

Thus the inclusion of the displacement current is needed to make Eq (1) valid for any surface S bounded by s

If one defines current as a transport of charge the term displacement current is certainly a misnomer when applied to a vacuum where no charges exist If however current is defined in terms of the magnetic fields it produces the expression is legitimate In a dielectric where an electric field produces a displacement of the negative charges with respect to the positive ones the name has meaning Maxwell had this sort of picture even for a vacuum where he postulated a polarizable ether See ETHER HYPOTHESIS [W R SM]

Bibliography See MAXWELL S EQUATIONS

Displacement pump

A pump that develops its action through the alternate filling and emptying of an enclosed volume

Reciprocating pumps Positive displacement reciprocating pumps have cylinders and plungers or pistons with an inlet valve which opens the cylinder to the inlet pipe during the suction stroke and an outlet valve which opens to the discharge pipe during the discharge stroke Reciprocating pumps may be power driven through a crank and connecting rod or equivalent mechanism or direct acting driven by steam or compressed air or gas

Figure 1 shows a small high speed plunger type power pump for high pressure service The three-throw crankshaft is carried in roller bearings at each end The manifolds below the suction valves and above the discharge valves connect the three

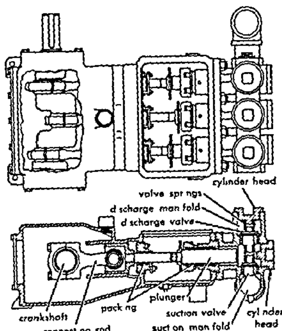


Fig 1 Horizontal triplex power pump

pumping cylinders to the suction and discharge piping

Power pumps are frequently built with one or two throw cranks and double acting liquid ends or with five seven or even nine cranks where smoother flow is desirable Power driven reciprocating pumps are highly efficient over a wide range of discharge pressures Except for some special designs with continuously variable stroke reciprocating power pumps deliver essentially constant capacity over their entire pressure range when driven at constant speed In some applications this is an advantage but in others it complicates the controls required

Direct acting steam pumps A reciprocating pump is readily driven by a reciprocating engine a steam or power piston at one end connects directly to a fluid piston or plunger at the other end Direct acting reciprocating pumps are simple flexible low speed machines which are low in efficiency unless the heat in the exhaust steam can be used for heating Steam pumps can be built for a wide range of pressure and capacity by varying the relative size of the steam piston and the liquid piston or plunger The delivery of a steam pump may be varied at will from zero to maximum simply by throttling the motive steam either manually or by automatic control Direct acting pumps are built as (1) simplex having one steam and one liquid cylinder and (2) duplex having two steam and two liquid cylinders side by side As indicated by Fig 2 each steam valve of a duplex pump is positively driven by the motion of the opposite piston rod by means of cranks and links In the case of a simplex pump to avoid stalling at low speed the valve linkage operates a small pilot valve which in turn controls a piston thrown main steam valve

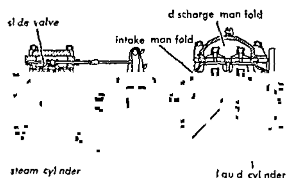


Fig 2 Duplex direct-acting steam-driven feedwater pump

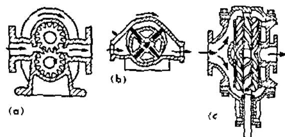


Fig 3 Rotary pumps (a) Gear (b) Sliding vane (c) Screw (*From L S Marks ed Mechanical Engineers Handbook 6th ed 1958*)

Reciprocating pumps are used for low to medium capacities and medium to highest pressures They are useful for low to medium viscosity fluids or high viscosity fluids at materially reduced speeds Specially fitted reciprocating pumps are used to pump fluids containing the more abrasive solids

Rotary pumps Another form of displacement pump consists of a fixed casing containing gears cams screws vanes, plungers or similar elements actuated by rotation of the drive shaft Most forms of rotary pumps are valveless and develop an almost steady flow rather than the pulsating flow of a reciprocating pump Three of the many types of rotary pumps are shown in Fig 3

Rotary pumps require very close clearances between rubbing surfaces for their continued volumetric efficiency Consequently they are most useful for pumping clean oils or other fluids having lubricating value and sufficient viscosity to prevent excessive leakage On petroleum oils of suitable viscosity rotary pumps are highly efficient at moderate pressure and speed while at reduced speed they can pump with lower efficiency the most viscous materials The increasing use of hydraulic actuation of machine tools and mechanisms such as power steering of automobiles has extended the use of rotary pumps and similar hydraulic motors

Vacuum pumps Although vacuum pumps actually function as compressors, displacement pumps are ... and condensing systems Sufficient liquid remains in the cylinder to fill the clearance volume and drive the air or gas out ahead of the liquid Certain types of rotary pumps are arranged to retain a quantity of sealing oil when operating as vacuum pumps

Air lift pumps In handling abrasive or corrosive waters or sludges where low efficiency is of secondary importance air lift pumps are used The pump consists of a drop pipe in a well with its lower end submerged and a second pipe which introduces compressed air near the bottom of the drop pipe The required submergence varies from about four times the distance from the water level to the surface for a low lift to an equal distance for a relatively high lift The mixture of air and water in the drop pipe is lighter than the water surrounding the pipe As a result the mixture of air and water is forced to the surface by the pressure of submergence *See* COMPRESSOR PUMP VACUUM PUMP [E F W]

Distillate fuel

In the distillation of petroleum fractions boiling above gasoline are used for a wide variety of fuels *broadly known as distillate fuels* The most important ones are kerosine furnace oils and diesel fuels Formerly heavy naphtha kerosine distillates of low octane number (known as engine distillates or tractor fuels) were used as fuels for low compression engines of farm tractors farm lighting units and small boats Such power units have been largely replaced by high compression gasoline engines or small diesels requiring different fuels *See also* DIESEL FUEL KEROSINE NAPHTHA PETROLEUM PRODUCTS [M SO]

Distillation

The process of producing a gas or vapor from a liquid or solid However the term sublimation is used ordinarily to describe the vaporization of a solid Heat is generally supplied to the liquid during the distillation although in special cases the latent heat required for the vaporization may be obtained from the internal energy of the liquid

The main purpose of distillation is either the separation of volatile components from nonvolatile materials or the separation of a mixture of volatile components The separation of volatile components from nonvolatile materials is carried out by a simple distillation in which the material is placed in a still and heated and the vapor removed and condensed Simple distillation is similar to the process of evaporation but it usually describes the operation in which the volatile material is a desired product whereas evaporation generally is applied to aqueous solutions of nonvolatile materials in which the nonvolatile material is the desired product Simple distillation is frequently used for high boiling organic compounds to prevent thermal degradation of the product the operation is usually carried out either at reduced pressure termed simple vacuum distillation or with the addition of

steam termed steam distillation *See* EVAPORATION, EVAPORATOR HEAT TRANSFER SUBLIMATION

Although simple distillation can be applied to mixtures of volatile components the separation obtained is usually not complete particularly if the components have boiling points that are close to each other To obtain greater separations in such cases *fractional distillation* is employed In this process the vapors from the still are permitted to come in contact with a portion of the condensate in a countercurrent or stepwise countercurrent operation Because of its lower operating and capital costs this countercurrent type of operation has completely replaced the multiple distillation condensation method formerly employed This process was originally developed in France for alcoholic beverages and it has been widely adopted for both laboratory and industrial operations because it is usually the most effective method of separating mixtures of miscible volatile liquids It is so effective and efficient that it is frequently employed to separate mixtures which are not normally liquids For example most industrial oxygen is produced by liquefying air and fractionally distilling the liquid air In this case the separation is so effective that not only high purity oxygen and nitrogen but also argon neon krypton xenon and other noble gases are recovered in commercial quantities The isotopes of hydrogen have been separated by fractional distillation at even lower temperatures In contrast is the high temperature fractional distillation of zinc and cadmium mixtures The temperature range between liquid hydrogen and liquid zinc includes the boiling points of most liquids and as a result fractional distillation has wide application

Vapor liquid equilibria Separation of two liquids is possible only when the composition of the vapor is different from that of the liquid from which it was produced The separation will be easier if there is a great difference between the composition of the vapor and that of the liquid but separations may be practical even when the difference is small The design of distillation equipment is usually based on vapor liquid equilibria compositions

The vapor liquid equilibrium data needed for distillation work are either obtained experimentally or estimated from physical chemistry relationships To obtain experimental data it is necessary to bring the vapor and liquid to equilibrium with each other Samples of each are then removed without altering the equilibrium and each phase is analyzed Such data are difficult and time consuming to obtain, consequently the amount of accurate vapor liquid equilibrium data available is very limited

Because of the difficulty of determining vapor liquid equilibrium data experimentally the calculation or estimation of such data is important The principles of physical chemistry form the basis of understanding the equilibria between liquid and vapor phases The rules of Raoult Dalton and Henry are of particular importance For nonideal solutions the equations of Margules and Van Laar have been used to estimate the activity coefficients and although they have worked well in some cases they have not been satisfactory for many other mixtures. The use of fugacities has been of great assistance in calculating the effect of high pressures on vapor liquid equilibria High pressures are frequently used so that low boiling materials can be fractionated using available water supplies as the coolant but the pressures must be lower than the critical pressure of the mixtures to obtain the desired operation In checking experimental vapor liquid equilibria data the Duhem equation has been particularly useful *See* ACTIVITY (THERMODYNAMICS) BOILING POINT, FUGACITY GAS LIQUID SOLUTION THERMODYNAMICS (CHEMICAL) VAPOR PRESSURE

A useful method of correlating vapor liquid data is the volatility of one component in a mixture relative to another This is called the relative volatility and is defined as

$$\alpha_{ab} = (y_a/x_a/y_b/x_b)$$

where α_{ab} is the relative volatility of component A to component B and y_a x_a y_b and x_b are the mole fractions in the vapor and liquid of A and B respectively If $\alpha_{ab} = 1.0$ the ratio of A to B is the same in the vapor and the liquid and no relative separation has been obtained between them If $\alpha_{ab} < 1.0$ the vapor contains less A relative to B than the liquid phase if $\alpha_{ab} > 1.0$ the reverse is true

Relative volatility changes slowly with temperature and may either increase or decrease with rise in temperature Because most distillations are carried out at essentially constant pressure (variable temperature) the low dependence of relative volatility on temperature increases its usefulness

For distillation purposes the data for binary mixtures are usually presented for constant total pressure on either a temperature composition diagram or a y x diagram that gives the mole fraction of one component in the vapor as a function of its mole fraction in the liquid (Figs 1 and 2)

The temperature-composition diagram gives more information but is not as easily used as a y x diagram and the latter is most often used Curve I is for benzene and toluene and is typical of systems that are approximated by Raoult's law The mole fraction of benzene in the vapor is always greater than its mole fraction in the liquid and the relative volatility of benzene to toluene shown in Fig 2 is relatively constant at about 2.4 Curve II is for a system in which the equilibrium curve crosses the $y = x$ line Thus at one point the vapor and liquid compositions are equal this composition is termed an azeotrope azeotropic mixture or constant boiling mixture because it will vaporize without any change in composition and therefore without any change in temperature during the evaporation Mixtures having less acetone than the azeotrope have a higher concentration of

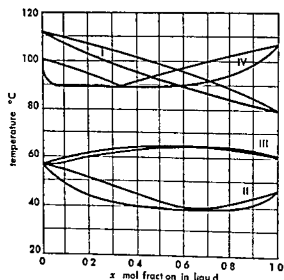


Fig 1 Temperature-composition diagram Curve I benzene toluene II acetone–carbon disulfide III acetone-chloroform IV isobutyl alcohol–water

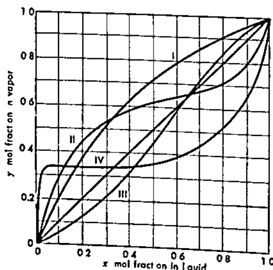


Fig 2 Composition-composition (y x) diagram Curve I benzene toluene II acetone–carbon disulfide III acetone-chloroform IV isobutyl alcohol–water

acetone in the vapor than in the liquid and the relative volatility of acetone to carbon disulfide is greater than 1 0 The relative volatility is 1 0 at the azeotropic composition and becomes less than 1 0 for mixtures containing more acetone In this latter region the concentration of acetone in the vapor is less than in the liquid The azeotrope or constant boiling mixture has a higher vapor pressure than either acetone or carbon disulfide and therefore at a given total pressure it boils at a lower temperature It is termed a minimum boiling azeotrope or a minimum constant boiling mixture A y x curve that crosses the $y = x$ line with a slope less than that of the $y = x$ line will have a minimum boiling azeotrope at the composition represented by the point of intersection

Curve III illustrates a mixture whose y x curve also crosses the $y = x$ diagonal but in this case the slope is greater than that of the $y = x$ line and the mixture corresponding to the intersection of the two curves will be a maximum boiling azeotrope This mixture will distill without change in composition and it will have a higher boiling point at a given pressure than either acetone or chloroform

Curve IV is similar to Curve II except that for a considerable region the vapor composition is constant This curve is characteristic of partially miscible liquid systems See EQUILIBRIUM PHASE

Simple distillation This operation can be carried out either as a continuous steady state operation or batchwise In the continuous system the liquid to be separated is added to the still as the feed at a steady rate a portion of it is vaporized by the heating coil the vapor produced is condensed as the distillate or overhead product and the unvaporized liquid is continually removed as the still or bottom product In such a system (Fig 3) the vapor product is usually of a composition close to the value in equilibrium with the liquid leaving the still For this type of distillation the material balance is

$$Vy + Lx = Fz$$

where V L F = moles per unit time of vapor unvaporized liquid and feed respectively and y x z = mole fractions of a component in the corresponding stream

Combined with the over all material balance this gives

$$\frac{V}{F} = \text{fraction vaporized} = \frac{z - x}{y - x}$$

which together with data relating y and x such as a y x equilibrium curve makes it possible to calculate y and x for any fraction vaporized

In a simple batch distillation the material to be separated is added to the still before the distillation is begun and no additional feed is added during the cycle As the distillation is continued the amount of liquid in the still decreases and the compositions of the vapor leaving and of the liquid remaining in the still continually change with time Rayleigh developed an equation for a binary distillation of this type If the relative volatility α is constant during the distillation the Rayleigh equation becomes

$$\ln \frac{W}{W_0} = \frac{1}{\alpha - 1} \ln \frac{x(1 - x_0)}{x_0(1 - x)} + \ln \frac{1 - x_0}{1 - x}$$

where W_0 is the original number of moles added to the still W is the number of moles remaining in the still x_0 is the original mole fraction and x is the mole fraction corresponding to W

Steam distillation In this operation steam is introduced directly into the liquid in the still The method is usually limited to those cases in which the solubility of the steam in the liquid is low at the operating temperature and pressure It is employed with relatively high boiling organic materi

als which would decompose if they were distilled directly at atmospheric pressure or with liquids that have such poor heat-transfer characteristics that excessive local overheating would result with indirect heating. By steam distillation a volatile material may be separated from nonvolatile impurities, or mixtures may be separated with results about equivalent to those obtained with simple distillation. Other gases or vapors could be used instead of steam, but steam usually is the most desirable from the viewpoint of cost and ease of recovering the vaporized materials. The heat required for vaporization must be supplied by indirect heat, or by the steam. For maximum steam economy, the temperature of the still should be as high as possible without undesirable thermal effects, and the total pressure should be as low as is consistent with the condensation of the vapor mixture.

When a simple distillation is carried out under a high vacuum, the rate at which molecules leave the surface rather than equilibria determines the composition of the vapor. Such an operation is termed molecular distillation. See CONDENSER, VAPOR; DISTILLATION, MOLECULAR; HEAT EXCHANGER.

Fractional distillation. In this operation, the vapor produced in the still is brought into contact with a portion of the condensate in a countercurrent or stepwise countercurrent system. The operation can be batch or continuous. The unit consists of a still to which a vertical column is attached (Fig. 4). This column is filled with some type of packing or plate construction which will permit the descending liquid added at the top to come into contact with the vapor rising from the still. The vapor from the top of the column is condensed and

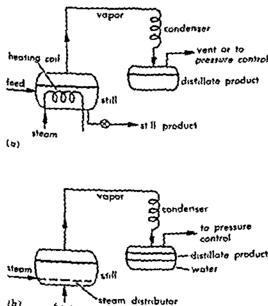


Fig. 3. (a) Continuous simple distillation. (b) Batch steam distillation.

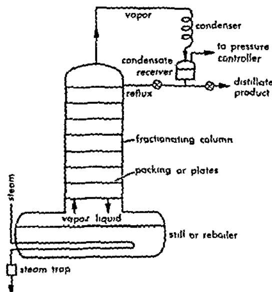


Fig. 4. Batch fractional distillation system.

a portion of it is removed as product; the remaining liquid is returned to the top of the column as liquid called reflux. The flow rates of liquid and vapor are adjusted so that at every high place within the column the liquid has a higher concentration of the more volatile components than corresponds to equilibrium with the vapor with which it is in contact. As a result the more volatile components pass from the liquid to the vapor and the less volatile components pass in the reverse direction. The vapor becomes progressively more enriched in the volatile components as it flows up the column to the condenser, and the liquid becomes more concentrated in the less volatile components as it flows down the column to the still.

The separation obtainable by such a system is much greater than that of a simple distillation; it depends on the relative volatility, the effectiveness of the transfer between the phases, and the ratio of the liquid to vapor in the column (or the fraction of the condensate returned as reflux as compared to that removed as product). The relative volatility is essentially fixed by the components involved, although it can be altered by changing the operating pressure. The effectiveness of the contacting device is determined by its ability to make a rapid exchange of the components between the vapor and the liquid; the exchange is mainly a function of interfacial area produced and the flow characteristics of the vapor and liquid. The interfacial area usually is proportional to the volume of the column, but increased height is more effective than increased cross-sectional area of the column in improving the efficiency of the interphase transfer because of the effects of vapor velocity and liquid distribution.

In a batch fractional distillation column, the overhead product is continually changing in composition and fractions of increasing boiling point

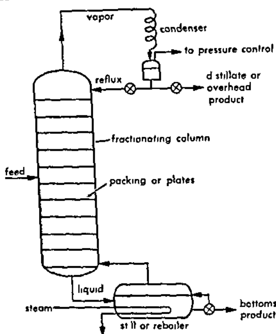


Fig 5 Continuous fractional distillation system

considerable labor is required to control the process It is most commonly used for laboratory fractionations and for small industrial operations When large quantities of materials are to be separated on a regular basis it is more economical to operate the system on a continuous basis (Fig 5) In this system the feed to be separated is added continuously at some position in the column vapor is introduced at the bottom and reflux is introduced at the top The vapor for the bottom of the column is usually produced by vaporizing a portion of the liquid from the column, the remainder of the liquid from the column is the bottoms product By such a system it is possible to obtain an overhead product that contains a high concentration of the more volatile components in the feed and a bottom product that contains a high concentration of the less volatile components High degrees of separation are obtainable and the factors that determine the effectiveness of the separation are the same as those described for the batch system The feed for such an operation can be liquid or vapor or a mixture of the two In cases where the bottom product is essentially water as in the fractionation of an ethanol water mixture it is possible to use steam directly as the vapor for the column instead of boiling the bottom liquid In cases where it is difficult to condense all of the overhead vapors for example petroleum fractions containing hydrogen or methane or where the overhead product is desired as a vapor it is possible to condense only that portion needed for reflux and to remove the rest as a vapor product

A very large number of different contacting arrangements has been used in the columns Laboratory columns and small industrial columns (1-2 ft in diameter or less) are frequently filled with packing which may be almost any small solid having a size from one tenth to one fiftieth that of the column diameter In laboratory columns many types of packings have been used and small rings, spheres, wire spirals, and special shaped and perforated metal pieces are common Packings may be made of metal, glass, ceramic, or carbon or any other material that can be suitably shaped and which will stand the operating conditions (Fig 6) In larger packed columns, it is customary to use rings or special shaped pieces, although packings such as coke lumps or bricks have been used For larger diameter columns, it is cheaper to use some type of plate arrangement instead of packing These plates may be perforated disks special grids, or more complicated arrangements such as bubble cap plates (Fig 7) Some of these are similar to packing in their action because the liquid flows down from plate to plate while the vapor passes up by it, but most of the plate units are designed to produce a bubbling type action. For example on a perforated plate there is liquid over the perforations through which the vapor bubbles as it flows up the column and the liquid flows down

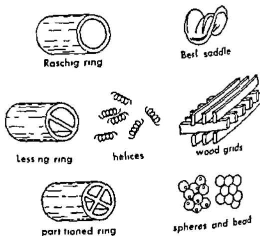


Fig 6 Typical column packings

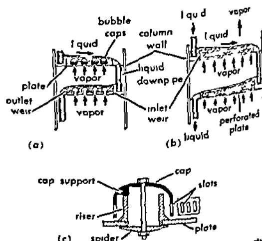


Fig 7 Typical plates (a) Schematic cross section through bubble-cap plate column (b) Schematic cross section through perforated plate column (c) Typical bubble cap

either through the perforations or through special pipes or channels arranged for this purpose

For large-diameter columns the plate construction is not only less expensive but is generally more effective for the interphase transfer because of liquid channeling encountered in large packed columns. The unsatisfactory liquid vapor contacting that results from channeling in packed towers can be improved by redistributing the liquid within the column *See* FRACTIONATING COLUMN

Design calculations for fractionating columns are usually made on the basis of the theoretical plate which was defined by E Sorel as a plate for which the average composition of the vapor leaving the plate was of a composition equal to the vapor in equilibrium with the liquid leaving the plate In such designs it is usually desirable to calculate two limiting cases as well as the actual operating conditions The effectiveness of the column increases as the ratio of the liquid reflux to overhead product called reflux ratio is increased the fewest theoretical plates will be required when this ratio is very large This limit is called total reflux The other limit is the lowest reflux ratio that will give the desired separation even if an infinite number of theoretical plates were used It is termed minimum reflux ratio Both of these limits require columns of infinite volume to obtain a finite product rate and are therefore not of practical design but they indicate the minimum number of theoretical plates and the minimum heat consumption that can be used for a given system

In the case of a steady state binary distillation for which the relative volatility is a constant M R Fenske has given an equation for the number of theoretical plates at total reflux for a column operating with a reboiler and a condenser that produces no separation between the reflux and distillate product

$$N + 1 = \frac{\ln [(x_a/x_b)_D/(x_a/x_b)_B]}{\ln \alpha_{ab}}$$

where N = number of theoretical plates x_a x_b = mole fractions of A and B respectively D and B refer respectively to overhead and bottom products and α_{ab} = relative volatility of A to B Figure 8 shows a plot of this relation

Many graphical and analytical methods for binary mixtures have been proposed for estimating the number of theoretical plates required as a function of reflux ratio and for total reflux and the minimum reflux ratio

Multicomponent mixtures are more complicated to study than binary systems and the necessary equilibrium and enthalpy data are usually not available Although a batch column can separate such a mixture into a number of fractions of relatively high purity the continuous column illustrated can give only two fractions In this case the overhead product could be relatively pure in the lowest boiling component but the bottoms would contain a mixture of the other components or the

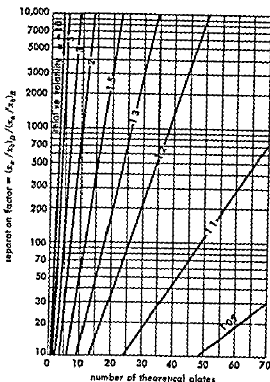


Fig 8 Number of theoretical plates at total reflux

bottoms could be relatively pure in the highest boiling component and the overhead product would contain a mixture of the more volatile materials A number of high purity fractions can be obtained from a multicomponent mixture with a continuous system by using a number of columns in series

Plate efficiency This term is used to express the relationship between the performance of an actual plate and a theoretical plate Many definitions have been given for plate efficiency but the simplest to use is the over all column efficiency which is the ratio of the number of theoretical plates required for a given separation divided by the actual number of plates needed This plate efficiency depends mainly on the characteristic of the mixture being separated and to a lesser degree on the design of the plate The most important characteristics of the system are the relative volatility and the viscosity of the liquid in the column

Although packings and other of the contacting methods do not give the definite stepwise arrangement obtained with perforated and bubble cap plates the analysis on the basis of theoretical plates is so convenient that this method is often employed and a height of packing that makes a separation equivalent to that of a theoretical plate is used This is called the height equivalent of a theoretical plate H E T P An alternate method of design uses the number of transfer units N T U and the height of a transfer unit H T U

As the relative volatility of the components approaches 1 0 (close boiling components or mixtures near an azeotropic composition) separation

by fractional distillation becomes difficult and large equipment and high heat consumption are required The relative volatility in such cases can frequently be changed by using a different total pressure which will make the separation easier but this is generally not very effective A more common technique is to add another component in the liquid phase that will so alter the volatilities that the separation of the desired components can be made more economically Depending on the characteristics of the added component this technique is termed azeotropic distillation or extractive distillation *See* MASS TRANSFER OPERATION TUBE STILL HEATER [E R G]

Bibliography Lord Rayleigh On the distillation of binary mixtures *Phil Mag* [6]4(23) 521–537 1902 C S Robinson and E R Gilliland *Elements of Fractional Distillation* 4th ed 1950

Distillation, molecular

A process by which substances are distilled in high vacuum at the lowest possible temperature and with the least damage to their composition The process has been used on a large scale to separate the relatively delicate vitamins A and E from fish liver and vegetable oils respectively The process differs from ordinary distillation in that the condensing surface is placed close enough to the evaporating liquid to catch the escaping molecules before they collide with one another or with mechanical obstructions in the still

Thus the molecular still is the limiting type of the class of open path stills which also includes short path stills By comparison an ordinary still (Fig 1) is a long path and restricted path still in which the molecular traffic of vapor in and out of surface is tremendous compared with the rate of escape through the pipe to the condenser Figure 2 suggests a molecular evaporator in which a slowly vaporizing surface of limited area is exposed to unlimited vacuum that is interstellar space and the molecule *a* will escape at the first try forever When a sufficiently cold condenser is brought to enshroud the area and to trap molecule *a* a state of molecular distillation is defined and the still is an open path molecular still (Fig 3) Should the condenser be brought close to the surface there will be formed a short path still which will be considered molecular as long as 90% of the particles are condensed at first emergence from the distilland Evidently the density or rate of molecular distillation can be increased as the path is shortened as in Fig 4 which shows a popular laboratory still

When the rate of evaporation is increased in the long path open path still a distillate atmosphere is created between evaporator and condenser which has an impressed velocity towards the latter At very high material rates the back diffusion of evaporated *molecules to the distilland* approaches a limit of 50% (Fig 5) so that an open path distillation need never be less than one-half as efficient as a molecular distillation

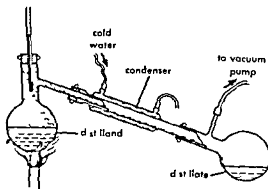


Fig 1 Long-path still

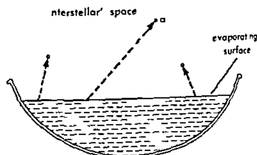


Fig 2 Molecular evaporator

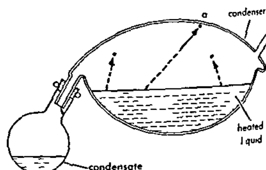


Fig 3 Open path molecular still

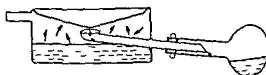


Fig 4 Laboratory molecular still

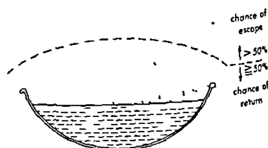


Fig 5 Long path open path still

The molecular still permits separation and purification at the lowest temperatures possible by distillation and thus with the least thermal hazard leading to decomposition Many unstable substances can be heated in the molecular still that would be utterly destroyed or even exploded by other stills The classic examples are the fat soluble vitamins which first came into commerce as concentrates through the open path still (vitamin A 1940 and vitamin E, 1942) Today this kind of still is used routinely in laboratories for the purification or preliminary separation of difficultly volatile mixtures particularly for heavy petroleum animal and vegetable fats steroid syntheses and industrially for separating mono di and triglycerides

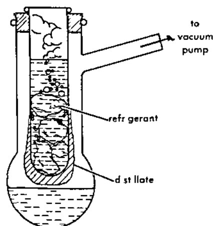


Fig 6 Cold finger still

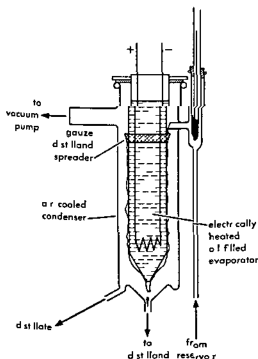


Fig 7 Falling film still

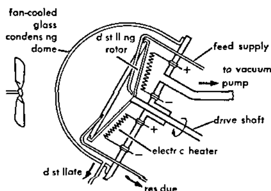


Fig 8 Centrifugal-disk still

To create an efficient open path still it is necessary not only to lower the density of the vapor and its obstruction but to agitate the surface of the distilland and spread it so thinly that there is very short exposure to high temperature These requirements have led progressively from the wide necked alembic to the open pot still the stirred pot still the rotating flask still the cold finger still (Fig 6) the falling film still (Fig 7) the centrifugal disk still (Fig 8) and the wiped film vertical still

The rate of molecular distillation of a pure substance is related to the vapor pressure and absolute temperature

$$Q \quad 0\,583\ p_\mu (M/T) E$$

where Q is the weight in grams evaporating per second per square meter p_μ is the vapor pressure measured in microns (10^{-3} mm Hg) The rate of evaporation Q_c of any constituent c is proportional to the partial pressure $p_{\ \mu}$ of c The factor E is the the evaporation coefficient which is near unity in the rapidly stirred or centrifugal still but may be less than 0 01 with a stagnant distilland A high molecular weight organic substance generally distills about 0 5–0 6 g/sec per square meter of surface *per micron* of vapor pressure At a manageable pressure of 408 μ from a medium sized industrial still this represents 15 30 kg of product per hour *See* DISTILLATION KINETIC THEORY OF MATTER MASS TRANSFER OPERATION [K C D H]

Distilled spirits

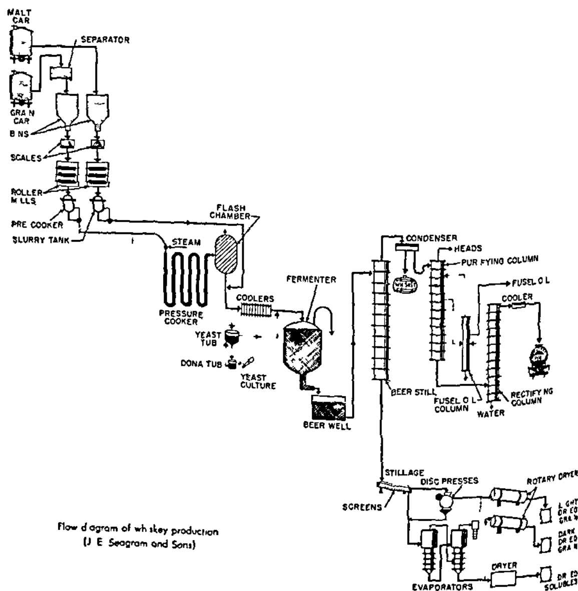
A spirit is a potable alcoholic beverage obtained by distilling an alcohol containing liquid and further treating the distillate to obtain a beverage of specific character

Basis of classification The various distilled spirits are classified according to (1) the raw material which has been fermented and subsequently distilled such as grain molasses and fruit, and (2) the further treatment given the distillate to add specific flavors and aroma These are referred to as compounded or flavored spirits Examples of beverages made from grain are whiskeys vodka, and gin Some types of vodka and gin are from potatoes Molasses are used for the ma

ture of rum and sugar cane juice for the Brazilian cachaça or pinga Fruits are used for the prepara tion of brandies, and agave juice for Mexican te quila The so-called compounded or flavored spirits are made from distilled spirits by the addition of sugar essential oils color herbs bitters or other ingredients Examples are English and Dutch gins various liqueurs or cordials absinthe aquavit and others

Fermentation When starchy materials are to be fermented it is necessary to convert the starch to fermentable sugars first This is done by the addi tion of amylolytic or diastatic enzymes from malt or certain fungi Once the starch has been con verted to sugar mostly maltose and glucose it can be fermented by yeast to alcohol and carbon diox ide When wine is distilled the fermentation is car ried out by various strains of the wine yeast *Sac charomyces cerevisiae* or its variety *ellipsoideus* With other sugar containing substrates special dis tillery yeasts may be employed usually also strains of *S cerevisiae*, or a natural fermentation by un controlled species of yeast is used

Distillation process This is done in stills of various design In a simple pot still very little frac tionation or rectification is accomplished When a still is equipped with a rectifying column frac tions of much higher purity can be collected The batch type distillation is becoming replaced more and more by continuous stills especially in the production of grain alcohol or neutral spirits Normally three types of fractions are separated A fraction with a low boiling point the "heads" is rich in aldehydes and esters The main fraction is ethyl alcohol and a higher boiling fraction is rich in so called fusel oils or higher alcohols The latter consists of D amyl and isoamylalcohol butyl and isobutylalcoh[illegible]
hol and amy
of fusel oil
pressed by t
100 proof spirit is a spirit containing 50% alcohol



Flow diagram of whiskey production
(J E Seagram and Sons)

by volume at a temperature of 60°F Each degree of proof is equal to ½% of alcohol

Types of distilled spirits The various types of distilled spirits are discussed below

Brandies Brandy is a spirit obtained from the distillation of wine or a fermented fruit juice usually after aging of the wine in wooden casks Cognac is a brandy distilled from wines made of grapes grown within the legal limits of Charente and Charente Inferieure Departments the Cognac region of France Armagnac is brandy made in the Department of Gers southeast of Bordeaux Both types of brandy are aged for many years in oak before bottling Before bottling various lots are blended the alcohol content is adjusted to 42-44% and if necessary coloring matter like caramel is added Brandies distilled from grape pomace of the wine press are called *eau de vie de marc* in France and *grappa* in Italy Spanish brandies are usually distilled from sherry wines and have a distinctive flavor quite different from cognac or armagnac American brandies are primarily products of California and have a flavor different from the European brandies Whereas in Europe pot stills are most common continuous stills are preferred in California Apple brandy is called applejack in the United States and is called calvados in France It is distilled from completely fermented apple juice and aged in oak barrels for 5–10 years It has a distinct apple flavor Other fruits from which brandy is made include black wild cherries (kirsch or kirschwasser) plums (slivovitz from Hungary or Rumania and quetsch or mirabelle from France) blackberries and apricots When stone fruits are used some of the stones are broken or crushed and a small amount of the oil is distilled over with the spirit giving the brandy a more or less pronounced bitter almond flavor

Whiskeys These distilled spirits are made by distilling fermented grain mashes and aging the distillate in wood usually oak Examples are Scotch whisky Irish whiskey Canadian whiskey rye whiskey and bourbon whiskey

Scotch whisky is made primarily from barley The barley is converted to malt by allowing it to sprout after which it is dried in a kiln over a peat fire The malt absorbs some of the smoke aroma which is carried over later with the spirit distilled from it The drying and roasting of the malt is to a large extent responsible fort the flavor of the whisky After the malt is made into a mash it is fermented distilled and aged in oak casks These are often casks in which sherry has been shipped A Scotch blended whisky may contain a certain percentage of grain whisky besides malt whisky

Irish whiskey is made from malt unmalted barley and other grains such as wheat rye and oats The malt used in Irish whiskey is not smoke cured as in Scotland and as a result the flavor of the resulting whiskey is different from Scotch whisky

Many types of whiskey are made in the United States Depending on the principal source of grain whiskeys are classified as bourbon (corn grain) rye whiskey (rye) and others Blended whiskeys are differentiated from straight whiskeys by a certain content of neutral spirits American whiskeys are aged for a number of years in new charred white oak barrels In all types of whiskeys discussed above the type and quality of the water used in the plant is an important factor in the quality of the finished product

Gins These consist essentially of a pure grade of alcohol which has been flavored with an extract of the juniper berry as the chief flavoring agent The flavor may be imparted by distillation of herbs (distilled gin) or by the addition of essential oils (a compounded gin) There are two principal types of gin English or London dry gin and Dutch gin (jenever)

English gin is made from pure neutral spirits which are redistilled in the presence of juniper berries and small amounts of other ingredients such as coriander seed cardamon seed orange peel anise seed cassia bark fennel and others English gin is not aged

Dutch gin is made of malt The flavoring ingredients are mixed directly in the mash after which it is distilled at a rather low proof Dutch gins are heavier in body and contain more fusel alcohols and other volatile compounds (congenerics) besides ethyl alcohol than does English gin Some

ent in cocktails

Rum The alcoholic distillate from fermented sugar cane juice or molasses is known as rum Cuban rum is light bodied and light colored The middle fraction of the distillation known as aguardiente is used for making rum The aguardiente is aged in uncharred oak barrels it is then decolorized filtered and supplied with some caramel to give it the proper color Occasionally some fruit aroma is added It is then aged for several more years Jamaica rum is heavier bodied and contains more congenerics (fusel alcohol esters and aldehydes) than Cuban rum Arak is a rum that comes from the island of Java It is a dry highly aromatic rum A natural fermentation of the molasses by various species of yeast also contributes to the flavor of this drink

Vodka A product originally produced in Russia but now popular in many countries vodka is usually made from wheat It is highly rectified during distillation and thus is a very pure neutral spirit without a pronounced taste It is not aged

Cachaça or pinga This is a Brazilian spirit made by distilling naturally fermented sugar cane juice in pot stills It is high in congenerics and sold at various degrees of proof It is usually not aged

Pulque and tequila Pulque and tequila are of Mexican origin A sweet sap is obtained from agave (century plant American aloe) by removing the growing bud from about 3 year ''

The sap (aguamiel or honey water)

and fermented by a natural fermentation and is then called pulque Tequila is obtained by distilling pulque

Liqueurs Liqueurs or cordials are alcoholic beverages prepared by combining a spirit usually brandy with certain flavorings and sugar In fruit liqueurs the color and flavor are obtained by an *infusion* process using the specified fruit and spirit Sugar is added after the extraction is complete Plant liqueurs are made by maceration of plant leaves seeds or roots with spirit and then distilling the product Sugar and coloring matter are added *after distillation* A large variety of different liqueurs are marketed *See* YEAST [E M M H J P]

Bibliography H J Grossman *Grossman's Guide to Wines Spirits and Beers* rev ed 1955

Distortion (electronic circuits)

Any undesired change in the waveform of an electric signal passing through a circuit including the *transmission medium* In the design of any electronic circuit one important problem is to modify the input signal in the required way without producing distortion beyond an acceptable degree Amplifier and loudspeaker systems are examples where maximum effort has been expended to produce a design for faithful amplification of speech and music input signals

There are four general types of distortion amplitude frequency phase and cross modulation The causes and effects of these types are discussed in this article

Amplitude distortion This is generally considered to mean distortion produced by a nonlinear relationship between the input and output amplitudes of a device One source of amplitude distortion is a vacuum tube The change in plate voltage of a vacuum tube is nearly proportional to the change in signal voltage only over a particular range of signal voltage amplitude As the amplitude is increased the change in plate voltage begins to depart from being proportional *See* VACUUM TUBE

To predict the amount of amplitude distortion that may be produced in a given case a power series representation for the characteristic of the tube is often used The relationship between plate voltage e_b and signal voltage e_s is expressed as

$$e_b = a_0 + a_1 e_s + a_2 e_s^2 + a e_s^3 +$$

The coefficients must be evaluated for each tube and the operating point used

The harmonic distortion is expressed in terms of percentages of the fundamental component of the plate voltage when the input signal e_s is a sinusoid If $A \sin \omega t$ is substituted for e_s in the terms of the power series and the powers of $\sin \omega t$ are reduced by trigonometric identities to a fundamental and harmonic components then the relative amplitudes of each of the harmonic components can be calculated for known values of the coefficients in the power series

In addition to distortion introduced by nonlinear tube characteristics distortion can be introduced in several other ways, all of which occur for large amplitude signals

Grid current Distortion is produced when the signal applied to the input is so large that the grid becomes positive with respect to the cathode The grid and cathode then form a diode and current flows through the grid circuit When this occurs the dynamic input impedance can change appreciably and the stage amplification is adversely affected

An additional form of distortion called blocking occurs in a resistance capacitance coupled amplifier stage when grid current flows in the following tube Under these conditions the coupling capacitor charges to nearly the plate supply voltage When the plate voltage drops and grid conduction stops the grid to cathode voltage of the following stage can become more negative than cutoff This condition persists until the capacitor discharges (through a large resistance and therefore at a much slower rate than it charged) sufficiently to unblock the following stage During the time that a stage is cut off there will be no amplification *See* VOLTAGE AMPLIFIER

Saturation When the input signal to a tube is large enough to drive the grid to-cathode voltage positive a further increase in the signal voltage produces little change in the plate voltage if the signal source output impedance is large compared to the dynamic grid to cathode resistance when the grid is conducting

Bottoming This form of distortion is produced by large input signals which drive the grid to-cathode voltage more negative than cutoff The plate voltage is then at its maximum positive value If the input signal goes more negative the plate voltage cannot increase and distortion results

Frequency distortion This form of distortion is an inherent feature of all amplifiers but can be minimized by proper design It occurs because the reactive elements and inherent reactances in the amplifier circuit do not allow the same amplification for all frequencies and therefore some components of a signal are amplified more than others Furthermore in the case of audio amplifiers the loudspeaker and enclosure characteristics affect the load presented to the amplifier in a manner which depends upon frequency For a discussion of frequency response *see* AMPLIFIER

The effects of frequency distortion may be considered in terms of an input signal composed of a fundamental and harmonic components such as a square wave If the amplifier gain is not a constant value for each frequency component the output will not be an amplified replica of the input

Phase distortion Like frequency distortion phase distortion is caused by the reactive elements in the circuit producing a phase shift that is not the same for each frequency component of the input signal It is possible for an input signal to have frequency components within the range of constant magnitude amplification but not within the range of constant phase shift for each component As a result the output is not an amplified rep-

tics of the input. Fortunately, the ear is more tolerant of phase distortion than frequency distortion. This fact simplifies the design of a high-fidelity amplifier system.

If the gain magnitude is constant with frequency while the phase shift changes uniformly with frequency, the output will be a replica of the input but delayed in time.

Cross modulation. Also called *intermodulation*, this effect is caused by nonlinear vacuum-tube characteristics. If two signals of different frequencies are applied to the input of a nonlinear vacuum-tube stage, the output will contain the fundamental and harmonic components of each signal, frequency components equal to the sum and difference of the input-signal frequencies, and sums and differences of the harmonics of the two input signals. Therefore, if the input is a signal composed of several frequencies the nonlinearity will produce new frequencies not integrally related to those of the input signal. The distortion, if bad enough, is generally more noticeable than harmonic distortion.

Reduction by feedback. Distortion caused by nonlinear vacuum-tube characteristics and by the frequency response of the amplifier can usually be reduced by the use of negative feedback. Amplitude distortion introduced in the last or next to last stages of a multistage amplifier can be reduced; distortion produced by the input stage cannot be reduced. Fortunately, in audio amplifiers distortion is generally produced in the last stage (the power-output stage).

The use of negative feedback in a properly designed amplifier will make the amplitude and phase shift more nearly constant over a wider frequency range than in an amplifier without feedback. This extended frequency range can be made to include for all practical purposes the frequency range of music and speech signals. *See* FEEDBACK CIRCUIT.

(H.F.K.)

Bibliography: J. D. Ryder, *Engineering Electronics*, 1957; S. Seely, *Electron-tube circuits*, 2d ed., 1958.

Distribution (probability)

The results of a series of independent trials, random variables, or errors often occur in fairly regular and predictable patterns. These patterns can be expressed mathematically, and the most important of them are called the binomial, normal, and Poisson distributions.

Binomial distribution. Consider n independent trials each of which results in success S or failure F with corresponding probabilities p and $q = 1 - p$. Denote by S_n the number of successes. Because there are $\binom{n}{k}$ possible ways to select k places for S's and $n - k$ places for F's, the probability distribution of the random variable S_n is given by $P\{S_n = k\} = \binom{n}{k} p^k q^{n-k}$, where $k = 0, 1, \ldots, n$. This is the binomial distribution. Its expectation is np, its variance npq. *See* PROBABILITY.

If one lets a random variable X_k equal 1 or 0 according as the kth trial results in S or F, then $S_n = X_1 + \cdots + X_n$. The *binomial distribution* can therefore be approximated by the normal distribution in accordance with the central limit theorem. This special case is known as the DeMoivre-Laplace theorem: putting

$$x_k = (k - np)/(npq)^{1/2},$$

it asserts that

$$P\{S_n = k\} \sim (2\pi)^{-1/2} \exp(-x_k^2/2)$$

$$P\{a < S_n < b\} \sim (2\pi)^{-1/2} \int_{x_a}^{x_b} \exp(-t^2/2)\, dt$$

with a percentage error tending to 0 as $n \to \infty$ provided x_a and x_b are restricted to a fixed interval. (With $p = q = \frac{1}{2}$ these formulas are useful for the evaluation of binomial coefficients and their sums. *See* BINOMIAL THEOREM.)

Normal distribution. The standard normal density is defined for $-\infty < x < \infty$ by

$$\varphi(x) = (2\pi)^{-1/2} e^{-x^2/2}$$

The standard normal distribution function or error function $\Phi(x)$ is its integral from $-\infty$ to x. The normal distribution function with mean m and variance s^2 is $\Phi((x - m)/s)$; its density is $(2\pi)^{-1/2} s^{-1} e^{-(x-m)^2/2s^2}$. As x goes from $-\infty$ to ∞ the function increases from 0 to 1. It plays an important role in many fields. In particular, $u(t,x;\xi) = (2\pi t)^{-1/2} e^{-(x-\xi)^2/2t}$ is the fundamental solution of the heat (or diffusion) equation $u_t = u_{xx}$ and represents the heat distribution on the x axis at time t caused by a unit heat source initially concentrated at $x = \xi$. Probabilistically, this represents the transition probabilities in the Wiener process.

A random variable whose distribution is normal is called normal or Gaussian. The sum of two independent Gaussian variables is again Gaussian, whereby the means and variances add. Analytically this means that the convolution of two normal distributions is again normal. This property characterizes the normal distributions among distributions with finite variances.

Central limit theorem and error theory. These are the best known and most important applications of the normal distribution. The nature of the central limit theorem is best seen from the simplest special case: if $X_1, X_2, \ldots$ are independent random variables having a common distribution with mean m and variance s^2, their sum $S_n = x_1 + \cdots + x_n$ has mean $\mu = nm$ and variance $\sigma^2 = ns^2$; the corresponding standardized variable $S_n^* = (S_n - \mu)\sigma^{-1}$ has a probability distribution $P\{S_n^* \leq x\}$ tending to $\Phi(x)$ as $n \to \infty$; in other words, the distribution of S_n gets close to $\Phi[(x - \mu)\sigma^{-1}]$.

The striking feature of this result is that an essential property of a sum of many independent components is independent of the nature of these components. The general central limit theorem shows this to be true under much wider conditions: the distribution of the sum S_n tends to normality even if the distributions of the components X_k vary with k provided only that the components are likely to be of the same order of magnitude (so that no component has individually a noticeable

effect on the sum) It is not even necessary that the X_k be independent

Many empirical quantities (for example the amount of water in a reservoir certain inherited characteristics such as height and the experimental error of physical measurements) represent the cumulative effect of many small components and the statistical fluctuations of such quantities may be expected to follow the normal distribution In particular under the authority of K F Gauss it has been assumed for a long time that experimental errors are approximately normally distributed Modern research has shown the limitations of this assertion Many distributions appear to the untrained eye as nearly normal but refined statistical tests prove that even the finest physical measurements depart considerably from normality and that assignable causes (that is large contributing components) can be discovered statistically This discovery shows the questionable character of the classical methods to predict the probable experimental error In industrial quality control the departures from normality are used efficiently to discover assignable causes and thus to spot coming trouble at an early stage *See* QUALITY CONTROL

Warning Much harm has been caused by the widespread misinterpretation of the limit theorems in probability and of the meaning of statistical equilibrium in stochastic processes The situation may be explained in terms of a coin tossing game in which H and T count 1 and -1 respectively Here $X_k = \pm 1$ with probability ½ and $m = 0$ $s = 1$ The operational meaning of the central limit theorem in this case is as follows Fix a large n Repeat the same game of n tossings independently many times One would then expect that in about 50% of the cases $S_n > 0$ in about 25% of the cases $S_n > 0.67n^{1/2}$ in about 16% of the cases $S_n > 2n^{1/2}$ and so on What the central limit does not say is that in one game about half of the sums $S_1 \quad S_n$ will be positive In fact the arcsine law shows the opposite to be true it is much more probable that all $S_i > 0$ than that there be equally many positive and negative ones

The multivariate normal distribution The above theory carries over without essential changes to n dimensions The n dimensional normal density is defined by $(2\pi)^{-n/2} D^{1/2} e^{-1/2 Q(x_1 \quad x_n)}$ where Q is a positive definite quadratic form with determinant D, the matrix of variances and covariances is the reciprocal of the matrix of Q If the n-dimensional joint distribution of the random variables $X_1 \quad , X_n$ is normal then each X_i is normal The converse assertion is false even though found in textbooks The multivariate normal distribution is important for stationary stochastic processes *See* STOCHASTIC PROCESS

Poisson distribution The Poisson distribution with parameter λ is the probability distribution of a random variable assuming the values 0 1 2 with probabilities $p_k(\lambda) = e^{-\lambda} \lambda^k / k!$ Both its expectation and variance equal λ This is one of the most important distributions it plays a basic role in the theory of stochastic processes and in many applications A full understanding of its character can be obtained only from a postulational derivation and from the consideration of its many generalizations However, much can be gained by the following elementary approach starting from the binomial distribution

Consider a large number n of independent trials each of which results in success or failure with probabilities p and $q = 1 - p$ Ordinarily interest is restricted to the case where p is very small but the expected number of successes $np = \lambda$ is of moderate size Typical examples may be obtained by considering centenarians color blind people or triplets in a large population, the defectives in an allotment of screws or fuses, or the wrong calls among all calls arriving during a day at a busy telephone The number of successes in the n trials is a random variable with the binomial distribution but under the present circumstances the binomial is close to the Poisson distribution and may be replaced by it In fact the probability of no success is $q^n = (1 - \lambda/n)^n$, which is close to $e^{-\lambda}$ the first term of the Poisson distribution Now the ratio of the kth to the $(k-1)$th term in the Poisson distribution is λ/k and in the binomial distribution $(n - k + 1)p/kq$ which tends to λ/k when $np = \lambda$ and $p \to 0$ The kth term of the binomial distribution approaches that of the Poisson distribution

This reasoning explains why the statistical fluctuations in the phenomena cited above follow approximately the Poisson distribution In other circumstances the Poisson distribution appears, not as an approximation but as the exact expression of a law of nature This is true in particular of processes where certain events such as radioactive disintegrations mutations power failures and accidents occur in time in such a way that (1) the probability that an event occurs during any given time interval of length dt is asymptotically λdt and (2) there is no interaction or aftereffect between nonoverlapping time intervals Under these assumptions a time interval of length t can be divided into $n = t/dt$ subintervals each representing a trial in which the probability of success is close to λdt a change to the limit $dt \to 0$ gives the Poisson expression $p_k(\lambda t)$ as exact probability for the occurrence of k events during a time interval of length t A similar argument applies to random distributions of points in space with t interpreted as volume, typical examples are stars flaws of material raisins in a cake and animal litters in a field For this reason "perfect randomness" of a chance aggregate of points in space or time is usually taken to mean that the fluctuations obey the Poisson law *See* ANALYSIS OF VARIANCE STATISTICS [W F]

Bibliography W Feller *An Introduction to Probability Theory and Its Applications* 2d ed Vol 1, 1957

Distributor

A rotary switch that directs the high voltage ignition current in the proper firing sequence to the various cylinders of an internal combustion en

gine In automotive practice, the distributor housing usually contains, in addition, apparatus for timing the ignition to occur when each piston is at optimum position in the cycle This apparatus includes a set of cam operated contacts called breaker points, the opening of which triggers the ignition pulse The timing of the breaker point opening is made earlier at high engine speeds by the centrifugal action of small weights that are driven by the breaker cam shaft (distributor shaft) Timing is also varied with engine load by the movement of a diaphragm exposed to the pressure in the engine intake manifold *See* IGNITION SYSTEM [A R R]

Disulfide

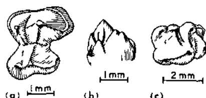
One of a group of organosulfur compounds RSSR′, that may be symmetrical (R = R′) or unsymmetrical (R and R′, different) They are of great biochemical interest, since the S—S link occurs in natural products such as cysteine α lipoic acid and insulin The cleavage of the S—S bond in proteins, for example, has important biological and industrial interests in the dehairing of hides, in the preparation of cysteine from wool in wave setting of hair, and in petroleum refining Disulfides are also of interest in the manufacture of polysulfide rubbers and polymers Higher linear polysulfides such as trisulfides and tetrasulfides are also known *See* AMINO ACIDS, INSULIN, ORGANOSULFUR COMPOUND, PETROLEUM PROCESSING, THIOFTHER [N K]

Division

Often called one of the fundamental operations of arithmetic and algebra It is a process of finding one of two factors of a number (or polynomial) when their product and one of the factors are given The symbol — now used mostly in elementary English and American arithmetics to denote division first appeared in print in an algebra by J H Rahn published in Zurich (1659) Division is more often symbolized by the double dot , the bar — or the solidus /, thus $x \ y$, $\frac{x}{y}$, or x/y indicates division of a number x by a number y Considered as an operation inverse to multiplication, x/y is a symbol denoting a number whose product with y is x Another way to base division upon multiplication is provided by the concept of the reciprocal of a number If y is any number (real or complex) other than 0 there is a number denoted by $1/y$ and called the reciprocal of y, whose product with y is 1 Then x/y is the symbol for the product of x and $1/y$ This view of division furnishes a means of extending the concept to objects other than real or complex numbers A whole number is divisible by 2 if its last digit is, and by 4 if the number formed by the last two digits is It is divisible by 3 or 9 according to whether the sum of its digits is, respectively and is divisible by 11 if the difference between the sum of the digits in the odd and the even places can be so divided *See* ADDITION, ALGEBRA, MULTIPLICATION, NUMBER SYSTEMS, NUMBER THEORY, SUBTRACTION [L M BL]

Docodonta

The docodonts, found in the Jurassic of North America and England, are probably the most primitive mammals known At present, all vertebrates in which the main jaw articulation is formed by the dentary bone of the jaw and the squamosal bone of the skull are classified as mammals The jaw of the ancestral reptiles consists of the dentary plus several other bones and the articulation is formed by the articular (jaw) and the quadrate (skull) bones In known docodonts the main jaw articulation is formed by the dentary and squamosal, however the articular and quadrate are believed to have formed a second jaw articulation In this way the docodonts bridge the gap between the Reptilia and the Mammalia



Upper and lower molars of *Docodon*, a Late Jurassic docodont (a) Occlusal view of an upper molar (b) Internal and (c) occlusal views of a lower molar (*After G G Simpson, 1929*)

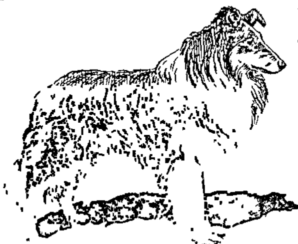
The dentition of the earliest docodont *Morganucodon*, consists of five upper and five lower incisors an upper and lower canine and six upper and seven lower postcanine teeth Unlike other mammals the alveolus of the last upper incisor is in the maxillary rather than the premaxillary bone The shoulder girdle of *Morganucodon* contains both precoracoid and coracoid bones and a branch of the trigeminal nerve leaves the skull through a foramen in the bone which surrounds the inner ear

Because of the similarity in structure of the shoulder girdle and inner ear, the most recent students of this order K A Kermack and F Mussett believe that the early docodonts were probably ancestors of the monotremes The unique characters of these mammals indicate that the docodonts were probably not ancestral to any of the other orders of mammals *See* MONOTREMATA FOSSILS [W A CL]

Dog

A carnivorous mammal *Canis familiaris*, of the family Canidae which has been domesticated by man since prehistoric times No other animal has been developed into so many types as the dog They hybridize freely with the coyote

Most dogs are pets, but they have practical use in hunting and transportation, in police war, and



The collie *Canis familiaris* height 20–40 in at shoulder (From E L Palmer, *Fieldbook of Natural History* McGraw Hill 1949)

treatment of diabetes is an outstanding example

Wild dogs may be quite destructive both to domestic animals notably sheep and to wildlife *See* CARNIVORA, COYOTE [J D B]

Dogfish

A primitive bony ganoid *Amia calva* sometimes called the bowfin occurs in fresh water in eastern North America Marine elasmobranchs such as *Squalus acanthias*, are also often called dogfish The dogfish is a dark brown cylindrical fish with a terminal mouth and a long dorsal fin Males have a prominent black spot at the upper base of the caudal fin ringed with orange-yellow Females lack the ring and the dark spot is obscure or lacking Females attain a length of 3 ft Males are somewhat smaller The dogfish is a voracious species preying upon other fishes and crayfish

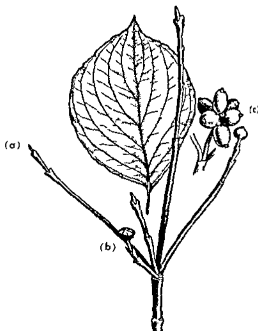


The dogfish or bowfin, *Amia calva* length of male to 20 in (From E L Palmer, *Fieldbook of Natural History*, McGraw Hill, 1949)

The female lays several thousand eggs in a nest similar to that of the sunfishes The eggs are tended by the [illegible] time a [illegible] the S [illegible] Young are often sold in [illegible] by bait dealers as mud minnows *See* AMIIFORMES [J D B]

Dogwood

A tree *Cornus florida*, also known as flowering dogwood which may reach a height of 40 ft and is found in the eastern half of the United States and in southern Ontario It has opposite simple deciduous leaves with entire margins When this tree is in full flower, the four large, white, notched bracts or petal like growths surrounding the small head of flowers give an ornamental effect that is unequaled by any native tree Pink, rose, and cream colored varieties are commonly planted The tree is tolerant of shade, so that at blossoming time usually in early May, the patches of white, even in dense woods, reveal its presence The wood is very hard and is used for roller skates, carpenters' planes, and other articles in which hardness is desired As a shade tree it is especially desirable for the modern ranch type house where small size is appropriate The Pacific dogwood *C nuttallii* which grows in Idaho and from southwestern British Columbia to southern California, is similar to



Flowering dogwood, *Cornus florida* (a) Vegetative bud (b) Flower bud (c) Fruit cluster (A H Graves *Illustrated Guide to Trees and Shrubs*, Harper 1956)

the eastern dogwood but has rounded bracts The Japanese dogwood, *C kousa*, is a similar small tree with pointed bracts that blooms in June Other shrubby species of dogwood are used as ornamentals *See* FOREST AND FORESTRY, TREE [A H G]

Dolomite

A common mineral with the ideal chemical composition $CaMg(CO_3)_2$ Commonly iron and manganese and more rarely cobalt, lead and zinc replace some of the magnesium Dolomite rock is formed of the mineral dolomite *See* DOLOMITE ROCK

Dolomite has hexagonal (rhombohedral) symmetry and single crystals may have curved faces It is colorless and transparent or white when pure It is more resistant to acid than calcite The hardness is 3½–4 on Mohs scale and the specific gravity is 2 85 Faceted crystals of dolomite are found in veins and cavities in limestones, dolomites and

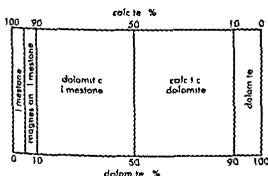
serpentines There are many localities where well formed dolomite crystals are found in Europe and South America In the United States large crystals have been reported from Roxbury and Chester Vermont and at Rochester Lockport and Niagara Falls New York

Dolomite can be readily synthesized at temperatures above approximately 200°C by heating a r mixture of magnesium and calcium carbo r portions under elevated

CARBONATE MINERALS]

Dolomite rock

A limestone whose carbonate fraction contains more than 50% of the mineral dolomite $CaMg(CO_3)_2$ To avoid confusion between the mineral and the rock R R Shrock has recommended the name dolostone for the rock but most geologists still use the name for both mineral and rock F J Pettijohn has suggested the names magnesian limestone for a dolomite content of 5–10% dolomitic limestone for 10–50% calcitic dolomite for 50–90% and dolomite for over 90% Most rocks called dolomite have more than 90% dolomite



Class ficat on of calcite-dolomite mixtures (*From F J Pett john Sed mentary Rocks Harper 2d ed 1957*)

Composition and texture Dolomites show gross textures similar to limestones but the finer details differ Most dolomite appears to be secondary rhombohedral crystals of dolomite that have grown in the rock after original limestone deposition usually have destroyed or obliterated many original structures and textures An original clastic texture may be revealed only by scattered quartz grains floating in a mosaic of dolomite crystals The original outlines of fossils or oolites may show as faint dust outlines in the dolomite but the original structure has been lost Fossil remains in dolomites tend to be casts and molds with poor preservation of structural details

The characteristic habit of dolomite crystals growing in a calcite matrix is rhombohedral The rhombohedrons are clearly formed after the calcite and in many cases have replaced it as is evidenced by the dolomite cutting across fossil shells and oolites The rhombohedrons may be zoned, successive growth stages having slightly different c position or amount of included foreign matter Complete dolomitization results in an interlocking mosaic of dolomite crystals in which because of interference between adjacent crystals rhombohedral faces may be lacking Partially dolomitized limestones tend to have a mottled appearance which results from the uneven distribution of dolomite crystals in the calcite matrix

In most of the rocks in which both chert and dolomite are found chert appears to have replaced dolomite and thus to have formed later In most rocks there is no firm evidence that dolomite has replaced chert Dolocasts hollow cavities in chert that have the shape of dolomite rhombohedra are common in the insoluble residues left from the acid digestion of dolomites and limestones These may be the result of dolomite replacing chert but may also represent selective replacement of a dolomitic limestone by chert or dolomite crystals being incorporated in a silica gel

Occurrence and origin Dolomite and dolomitic limestone are known from rocks of all ages but are more common in older rocks, particularly the Paleozoic Dolomite is most often found in association with limestone with which it may be interbedded or laterally gradational Some dolomitized zones do not follow bedding planes and are thought to be controlled by faults or folds Dolomitization of limestones may be highly selective for example the cores of the Silurian reefs of Illinois Indiana and Wisconsin are dolomite whereas the reef flank material may be only partially dolomitic Dolomite is also found in association with the evaporite rocks gypsum anhydrite and rock salt Evaporite dolomite is very fine grained and showing no evidence of replacement origin is thought to represent a primary precipitate from very saline waters See EVAPORITE (SALINE)

Most dolomites are replaced limestones as is evidenced by crystals of dolomite cutting across original textures by structural control of some dolomitization and by dolomitization cutting across stratigraphic boundaries The replacement of calcite by dolomite has been volume for volume not molecule for molecule for the latter would lead to an overall reduction in volume of 12% and dolomites as a group do not have significantly higher porosities than limestones The replacement is the product of the reaction between Mg^{+} ions in interstitial waters and calcite to form the double salt dolomite An alternative explanation is that there has been a transformation from a magnesian calcite (calcite that has magnesium replacing some calcium in the calcite lattice) to a thermodynamically more stable association of magnesium free calcite and dolomite Many of the carbonate secreting invertebrates incorporate magnesium in their calcite shells This may be the source of some dolomite But the reaction with magnesium in ground waters must be of major importance in producing the pure dolomites for there is not enough magnesium in the magnesian calcites to account for all dolomite See LIMESTONE, SEDIMENTARY ROCKS. [R

Dolphin

Any of several small to moderate sized whales of the family Delphinidae occurring in salt water throughout the world This family also contains the killer whales and blackfish Dolphins are similar to

The bottle nosed dolph n *Turs ops truncatus* length to 12 ft (From P Martin Duncan ed Cassell's Natural H story Cassell & Company Ltd London)

the related porpoises and like the latter have many small teeth on each jaw However dolphins have cone shaped teeth in contrast to the spade shaped teeth of the porpoises Dolphins may be as long as 20 ft although most species are 14 ft or less in length They vary greatly in size color and head shape Their food consists primarily of fishes and cuttlefish There are 10 species of dolphin along the coasts of the United States in addition to the killer and false killer whales and 3 species of blackfish

The common name dolphin is also applied to certain brilliantly colored tropical marine fish of the genus *Coryphaena* See CETACEA WHALE [J D B]

Domain (crystallography)

A region in a solid within which elementary atomic or molecular magnetic or electric moments are aligned

Ferromagnetic domains are regions of aligned magnetic moments Each domain may be thought of as a tiny magnet ...

A ferron ...

domains

the orig ...

magnetic domains see FERROMAGNETISM for the effect of domains on magnetization see MAGNETIZATION

Ferroelectric domains are electrical analogs of ferromagnetic domains See FERROELECTRICS

[E A F K]

Dominance, recessiveness, blending

... forms of a gene controlling that trait are present in the same organism For example in rabbits a gene designated *Y* determines the white fat usually found in these animals an alternative form of the gene *y* determines yellow fat Most white fatted rabbits have the genetic constitution *YY* yellow fatted rabbits have the genetic constitution *yy* Hybrids between the ... have one of each ...

be designat ...

fatted ... be dominant over *y* because when present together *Y* seems to override the effect of its alternative gene, which is designated as recessive In general in the notation of genetics dominant genes are designated by capital letters, their recessive alternatives by corresponding small letters White fat in rabbits might be called a dominant trait yellow fat a recessive trait See GENE, GENETICS

Dominance and recessiveness are not always complete Genes controlling the antigens which are the basis for human blood types produce these antigens irrespective of the presence of alternative forms of these genes (see BLOOD GROUPS) In the garden flower four o'clock, red flowered plants have the genetic constitution *RR*, white flowered plants carry two of an alternative gene and are *rr* However plants of the genetic constitution *Rr* produce pink flowers Because neither gene *R* nor its alternative *r* can be called dominant relationships of this kind are often called incomplete dominance ...

type alternatives Unfortunately the term blending sometimes carries with it the implication of an outmoded conception of inheritance according to which the heredity of an individual depended upon a mixing of the parents' bloods In fact and to the contrary the general rule is that the elements of heredity (genes) retain their discreteness so that a hybrid of the type indicated above (*Rr*) can transmit either an *R* or an *r* gene to a progeny individual the alternative forms of the gene being unchanged and undiluted by their association in the hybrid

Blending inheritance is a term that has also been applied to the hereditary manifestations of characteristics such as height weight and intelligence when various manifestations of the characteristics do not fall readily into simple or discrete classes In these instances the characteristics are influenced by many different genes that are members of different gene pairs With reference to such characteristics simple alternatives of dominance and recessiveness do not apply [A M S]

Donkey

A tough small member of the family Equidae *Equus asinus* also known as the ass or burro It is native to Africa and Asia and has been dome ...

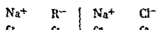
The donkey *Equus asinus* height to 41 ft (From P Martin Duncan ed Cassell's Natural H story Cassell & Company Ltd London)

cated for centuries The sparsely haired tail scraggly mane large ears and gray make this short haired small footed animal the object of much ridicule Nevertheless in warm climates it is the most satisfactory pack animal being durable and sure footed in rocks and along mountain trails Its slow deliberate gait and stubborn disposition are equally famous Males are known as jacks or jack asses, females are called jennets or jennys *See* MULE PERISSODACTYLA [J D B]

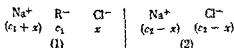
Donnan equilibrium

The particular equilibrium set up when two co existing phases are subject to the restriction that one or more of the ionic components cannot pass from one phase into the other This equilibrium was first recognized by F G Donnan in 1911 Commonly the restriction is caused by a membrane which is permeable to the solvent and small ions but impermeable to colloidal ions or charged particles of colloidal size The presence of a membrane is not essential since the restriction of movement on the charged colloid can be provided by a centrifugal or gravitational field or by gel coherence

An immediate consequence of such a restriction in a system is the uneven distribution of diffusible ions at equilibrium This is apparent in the following example Let the initial state of a system be a solution of the ions Na^+ and R^- of concentration c_1 separated from a solution of sodium chloride of concentration c_2 by a membrane freely permeable to all but the R^- ions



Then at equilibrium a certain concentration of chloride ions x will have diffused through the membrane accompanied by the same number of sodium ions in order to preserve electrical neutrality on both sides of the membrane and the final equilibrium will be



It can be shown thermodynamically that at equilibrium the product of the concentrations or more strictly the activities of the sodium and chloride ions shall be the same on both sides of the membrane

Hence

$$[Na^+]_1[Cl^-]_1 = [Na^+]_2[Cl^-]_2$$

or

$$(c_1 + x)x = (c_2 - x)^2$$

Obviously the diffusion of the chloride ions (and an equal number of sodium ions) through the membrane has been hindered by the presence of the nondiffusible ion R^- Calculations based on this equation which are confirmed by experimental measurements show that sodium chloride is almost completely prevented from diffusing through the membrane if it is present in small concentration relative to the concentration of the nondiffusible ion R^- As the relative concentration of sodium chloride is increased more of it diffuses through the membrane Finally when the salt concentration is very high relative to that of R^- an even distribution of sodium and chloride ions on either side of the membrane is approached A similar equilibrium is also attained when the diffusible salt has no ion common to the colloidal electrolyte

An important example of Donnan equilibrium is the dialysis of a solution of a colloidal electrolyte against pure water Sodium ions from the colloidal $R^- Na^+$ will diffuse through the membrane and be replaced by an equivalent number of hydrogen ions This phenomenon is called membrane hydrolysis and is helpful in explaining certain membrane equilibria in biological cells and tissues Obviously if the water is renewed continuously complete hydrolysis will ultimately ensue

Two important consequences arise from the Donnan equilibrium The first is that the observed osmotic pressure that is the difference in hydrostatic pressure on the two sides of the membrane will always exceed that of R^- except when a large excess of salt is added An illustration of this effect is the behavior of ionic gels (for example protein gels) when immersed in water Ionic groups attached to the structure of the gel cannot diffuse out into the surrounding solution and osmosis causes swelling of the gel The gel will obey the Donnan equilibrium the swelling is found to be reduced by the addition of salts

The second consequence of the Donnan distribution is that a potential difference E is set up at the membrane It is given by the equation

$$E = \frac{RT}{F} \ln \frac{[Na^+]_1}{[Na^+]_2} = \frac{RT}{F} \ln \frac{[Cl^-]_2}{[Cl^-]_1}$$

where R is the gas constant T is the absolute temperature and F is Faraday's constant This is the origin of the difference in potential between a suspension and its intermicellar liquid In soil chemistry this is known as the suspension effect *See* COLLOID DIALYSIS, ION PERMEABLE MEMBRANE [C S M , W O M]

Donor atom

An impurity atom in a semiconductor which can contribute or donate one or more electrons to the crystal by becoming ionized and positively charged For example an atom of column V of the periodic table substituting for a regular atom of a germanium or silicon crystal is a donor because it has one more valence electrons which can be detached and added to the crystal Donor atoms tend to increase the number of conduction electrons in the semiconductor The ionization energy of a donor atom is the energy required to dissociate the electron from the atom and put it in the conduction band of the crystal *See* ACCEPTOR ATOM SEMICONDUCTOR [H Y

Doppler effect

A change in the observed frequency of sound light or other waves caused by motion of the source or of the observer A familiar example for sound waves is the increase (decrease) in pitch of a train whistle as the train approaches (passes) The opti

one another more light pulses are received in a given time interval and the color emitted from the star appears to be shifted toward the violet end of the spectrum When the distance between the earth and the star is increasing the observed light is shifted toward the red end of the spectrum The color shifts of remote galaxies are taken as evi dence that the universe is expanding *See* RED SHIFT

In astronomy color differences between the ap proaching sides and reced

to compu

another ir

see DOPPL

Acoustical Doppler effect Acoustical observa tions of a moving source emitting sound at a con stant frequency make its pitch appear greater when the source is approaching the listener and smaller when the source to listener distance is increasing The effect is based on the fact that the listener perceives as frequency the number of sound waves arriving per second

The acoustical Doppler effect deals with cases of relative motion between the listener and the source and includes the effects of motion of the medium itself relative to both the source and the listener The wave velocity u of the sound in the medium is a property of the medium and its value is referred to that medium The wavelength λ frequency f, and velocity u are related in wave propagation by the equation $u = f\lambda$

A distinction needs to be drawn between the case in which the source moves relative to the listener fixed in the medium and the case in which the listener moves with respect to the source fixed in the medium

In the first case if the source moves toward the fixed observer with a velocity v_S waves emitted with a frequency f_S appear to have their wave length shortened in the ratio $(u - v_S)/u$ because of a crowding of the waves (Fig 1) which how ever still arrive at the listener with a velocity u

In the second case if the listener moves toward the fixed source the waves appear to him to arrive with a velocity $(u + v_L)$ The wavelength of the sound in the medium is unchanged in this case and is equal to that measured when both the listener and the source are fixed in the medium

Consider now the effect of the velocity of the medium relative to the listener and the source If v_M is the component of this velocity taken positive in the direction from the listener to the source and if v_L and v_S are the velocity components along the line joining the listener to the source and are now taken to be positive in the direction from the lis tener to the source then the general equation relat ing the observed frequency f_L and the source fre quency f_S is

$$\frac{f_L}{u + v_L - v_M} = \frac{f_S}{u + v_S - v_M} \qquad (1)$$

Optical Doppler effect This phenomenon seems at first to be analogous to the acoustical Doppler effect but the causes detailed effects and explana tion of the optical Doppler effect are fundamentally different and result from the relativistic behavior of light *See* LIGHT, RELATIVITY

Differences between the two effects Three funda mental differences exist between the acoustical and the optical Doppler effects

1 The optical frequency change does not de pend upon whether it is the source or the observer that is moving with respect to the other whereas the acoustical frequency change is different in the two cases

2 No effect is observable in the acoustical case when the source or the observer, moves at right angles to the line connecting the source and the observer An optical frequency change is observa ble under such conditions

3 The motion of the medium through which the waves are propagated does not affect the observed optical frequency whereas it does affect the ob served acoustical frequency

Light source in motion The mathematical ex pressions of the observable effects involving light and other electromagnetic waves are arrived at by noting that the propagation of a given plane wave must be described by the same law in the source frame and the observer frame according to the relativistic principle of equivalence Accordingly the equation of propagation of the plane wave written for the source frame of coordinates is transformed to the observer frame of coordinates with the help of the well known Lorentz transfor mations and the relevant factors on the two sides of the resulting equation identifying the descrip tions in the two frames are identified The result is expressed in Eqs (2) and (3)

$$f_O = \frac{f_S\sqrt{1 - (v^2/c^2)}}{1 - (v/c)\cos\theta_O} \qquad (2)$$

$$\cos\theta_O = \frac{\cos\theta_S + (v/c)}{1 + (v/c)\cos\theta_S} \qquad (3)$$

relating the frequency f_O and angle θ_O measured in the observer frame to the frequency f_S and angle θ_S that would be measured in the source frame under the conditions in which the source frame is measured (in the observer frame) to move with a velocity v relative to the observer frame In these equations c is the velocity of light in free space

Examination of the frequency relation shows that it incorporates two factors a purely relativistic direction independent factor $f_O \sim f_S\sqrt{1 - (v^2/c^2)}$ according to which the observed frequency will be smaller than the source frequency regardless of the

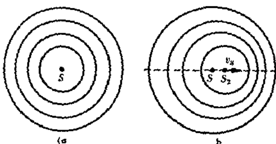


Fig 1 Spherical waves from a point source (a) At rest (b) in motion

apparent direction of motion of the source (transverse Doppler effect) and a direction dependent factor $f_O \sim f_S/[1 - (v/c) \cos \theta_O]$ showing a further dependence on the direction of relative motion Like the acoustical Doppler effect this factor is understandable on the basis of classical arguments

The part involving the direction of relative velocity (radial Doppler effect) can be derived by counting as the observed frequency f_O the number of waves arriving in a time interval dt_O corresponding to the difference in the times of arrival of a first wave and of a last wave traveling with a velocity c toward the observer The waves are emitted at a frequency f_S by a source traveling with a velocity v at an angle θ_O with respect to the observer During a given time interval dt_S the source emits a total of $f_S\, dt_S$ waves The relativistic velocity dependent part is then included by noting that the source frequency will appear to be $f\sqrt{1-(v/c)^2}$ according to the special theory of relativity

H E Ives and G R Stilwell (1938) skeptical as to the conclusions of the special theory of relativity set out to verify the velocity dependent part of the frequency shift (transverse Doppler effect) observed at zero angle ($\theta_O = 0$) By measuring the wavelengths of the H_β line in the direction of motion ($\theta_O = 90°$) of hydrogen canal rays at 18 000 volts (Fig 2) and in the opposite direction ($\theta_O = -90°$) for which the frequencies are respectively $f_{O1} = f_S\sqrt{1-\beta^2}/(1-\beta)$ and $f_{O2} = f_S\sqrt{1-\beta^2}/(1+\beta)$ where $\beta = v/c$ they determined the average $f_O = (f_{O1} + f_{O2})/2 = f_S/(\sqrt{1-\beta^2})$ This result was found to be in accord with the theoretical value $f_O = f_S/\sqrt{1-\beta^2}$ thus providing a direct proof of the dilatation of time according to which the observer thinks that the source period T_S is $T_O/\sqrt{1-\beta^2}$ and therefore greater than the observer period T_O [C W ST]

Bibliography R W Ditchburn *Light* 1953 A Einstein et al *The Principle of Relativity* 1926 M von Laue Relativitatstheorie Doppler und andere spektrale Verschiebungseffekte *Naturwissenschaften* 41(2) 25–29 1954 W C Michels, Phase shifts and the Doppler effect, *Am J Phys.,* 24(2) 51 53 1956 L Page *Introduction to Theoretical Physics* 3d ed 1952 A J W Sommerfeld *Lectures on Theoretical Physics* vol 4 1954 F W Sears and M W Zemansky *University Physics* 2d ed 1955

Doppler radar

A radar system used to measure the relative velocity of the system and the radar target The operation of these systems is based on the fact that the Doppler frequency shift in the target echo is proportional to the radial component of target velocity *See* DOPPLER EFFECT

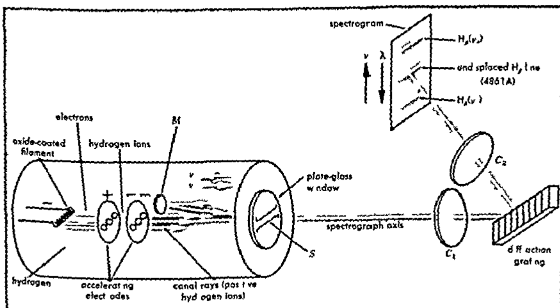


Fig 2 Ives and Stilwell experiment C_1 and C_2 collimating and camera lenses S entrance slit of spectrograph M concave mirror focused on slit and to observe light emitted from hydrogen ions moving away from slit (G W

Air borne systems are used to determine the velocity of the vehicle relative to the earth for such purposes as navigation bombing and aerial map ping or relative to another vehicle for fire control or other purposes Ground or ship equipment is used to determine the velocity of vehicular targets for fire control remote guidance intercept control traffic control and other uses

A Doppler system (Fig 1) consists of at least the following elements transmitter antenna as sembly receiver Doppler frequency measuring de vice and output signal generators or displays

The Doppler frequency shift Δf is an extremely small fraction of the transmitter frequency f It is given by (see Fig 2)

$$\Delta f = \frac{2Vf}{C} \cos \gamma$$

where V is the relative speed C is the speed of signal propagation and γ is the angle between the velocity and the direction of propagation The only practical way to measure it is by adding the echo signal to a reference signal derived from the trans mitter and observing the difference or beat fre quency Some means of obtaining coherent detec tion is required

Practical techniques have been devised for ob taining the requisite coherence in continuous wave pulsed and frequency modulated transmission sys tems For general discussion of these techniq ies see CONTINUOUS WAVE RADAR RADAR For a discus sion of techniques similar to those employed in Doppler radar see MOVING-TARGET INDICATION

Doppler navigation radar is a type of air borne Doppler radar system for determining aircraft ve locity relative to the earth s surface

The most important design considerations unique to Doppler navigation radar relate to (a) the number of beams (b) antenna stabilization (c) antenna design and (d) modulation and de tection techniques

The signal from a single beam can provide only the velocity component in the direction of that beam Complete velocity determination requ res, therefore the use of at least three beams Mo t sys tems use three or for symmetry four beams

To relate the beam directions and hence the measured velocity to an earth-oriented coordinate system a vertical reference must be provided The antennas may be fixed to the aircraft The Doppler frequencies and the vertical data (such as roll and pitch) are fed to a computer Its outputs are elec trical signals or displays representing the compo nents of velocity

Alternatively the antenna assembly is stabil zed in pitch and sometimes in roll as well utilizing vertical reference data The assembly is rotated in azimuth until Doppler frequencies from right and left directed beams are equal An axis of symme ry is then aligned in the direction of the horizontal component of velocity The outputs are ground speed (and in some cases rate of climb) deri ed directly from Doppler measurements and drift an gle from antenna orientation

Various types of antennas have found use in Doppler navigation radar Paraboloid and micro wave lens antennas are generally used in fixed an tenna systems Linear or planar arrays are gener ally used in stabilized antenna systems Since both volume and radome cutout area should be small various techniques are employed to enable each an tenna to form (simultaneously or sequentially) more than one beam Pencil beams of 3-5° width are used as are beams of elongated cross sect on The beams are directed 20–30° from the vertical Larger values result in insufficient echo power over water

Continuous-wave systems are coherent and are theoretically the most efficient in use The ch ef difficulty is control of leakage of spurious signals from transmitter to receiver

Pulsing enables the receiver to be rendered in sensitive during transmission thereby avoid ng leakage signals Coherence is achieved either by driving a transmitting power amplifier from a CW oscillator or by mixing at the detector a pair of

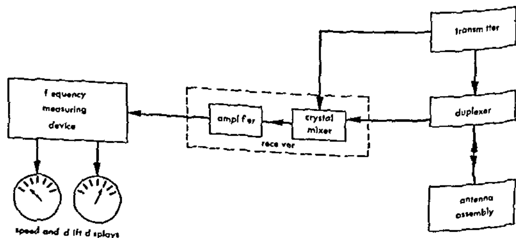


Fig 1 Block d agram of a s mple Doppler system

pulse echoes received over different propagation paths Pulsed systems are subject to altitude hole effects, which are associated with partial loss of return signal at times which are multiples of the transmitter modulation period

Some Doppler systems employ sinusoidal frequency modulation A sideband of the detected

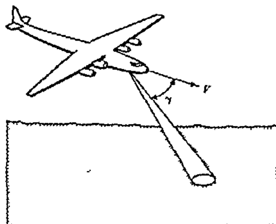


Fig 2 Basic Doppler frequency measurement geometry

beat between echo and transmitter signal is used Modulation index and rate and the sideband order are chosen such that echoes from nearby objects are rejected while those from distant objects are accepted Leakage noise is reduced at the expense of lowered efficiency Altitude hole effects occur

An example of a system (Fig 3) is the first one produced in quantity (1954) It bears the nomenclature AN/APN 81 This system's velocity accuracy is such that when averaged over 20 nautical miles of travel the ground speed error (standard deviation) is 0 04% and the drift angle error is 0 09 degrees This radar is used with a heading reference and a navigation computer to form a complete self contained automatic navigation system in which velocity components are integrated continuously to obtain present position One such system (AN/APN 66) shows an over land position error (circular probable error) of about $11/\sqrt{D}\%$ of distance traveled where D is the distance in nautical miles [F B B]

Bibliography E J Barlow, Doppler radar, *Proc IRE* 37(4) 340–355 1949 F B Berger, The nature of Doppler velocity measurement, *IRE Trans*, ANE 4(3) 103–112 1957, F B Berger, The design of airborne Doppler velocity measuring systems, *IRE Trans*, ANE-4(4) 157–175, 1957, M A

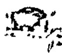


control panel

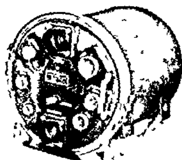




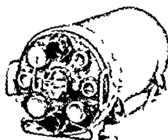
ground speed indicators



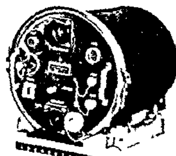
vertical gyro



frequency-tracker amplifier



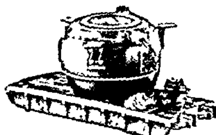
wind memory amplifier



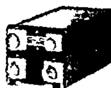
receiver transmitter



interconnection box



antenna



frequency tracker computer

Fig 3 Doppler radar set AN/APN 81 (G cision Laboratory, Inc)

Condie Basic design considerations—automatic navigator AN APN 67 *IRE Trans* ANE 4(4) 197 201 1957 W R Fried Principles and performance analysis of Doppler navigation systems *IRE Trans*, ANE 4(4) 176–196 1957 I R Mall ing Radio Doppler effect for aircraft speed measurements *Proc IRE* 35(11) 1357 1360 1947 F A McMahon The AN/APN 81 Doppler navigation system *IRE Trans* ANE 4(4) 202–211 1957

Dopplerite

A naturally occurring gel of humic acids found in peat bogs or where an aqueous extract from a low rank coal can collect Dopplerite is soluble in alkali contains organically bound calcium iron or magnesium and has an ash content of about 5% On an ash free basis it consists on the average of 55 5% carbon 5 5% hydrogen 36 0% oxygen 2 0% nitrogen The composition of the material is usually nearly identical to that of the coal from which it is derived On dehydration dopplerite becomes a black brittle solid with conchoidal fracture *See* Coal Peat

[I A B]

Dorylaimoidea

One of the most common groups of soil and fresh water inhabiting nematodes with the taxonomic status of an order or superfamily This group also includes two known marine species World wide distribution of many species attests to their ancient lineage among the Nematoda The almost endle [illegible]

[illegible] living organisms are found Predatory species feed on earthworms nematodes and other microorganisms Mycophagous forms browse on the mycelia of fungi whereas the intestines of other species are filled with algae and fragments of mosses Many are numbered among the most economically important plant parasites especially those of the subfamily Longidorinae One of these the American dagger nematode *Xiphinema americanum* Cobb is perhaps the most widespread and destructive plant parasitic nematode in the United States where it attacks and destroys rootlets of plants varying from wheat in Nebraska to orchards in New York and forest trees in Wisconsin It is also known in Israel and Ceylon Investigations have not been made of the life histories and food habits of hundreds of species of dorylaims *See* Nematoda

[G C T]

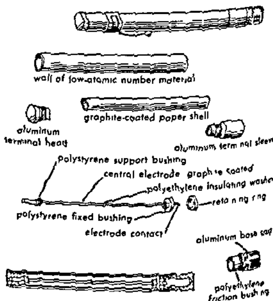
Dosimeter

A meter used for measuring the dose of ionizing radiation received by the person wearing it There are many types of dosimeters but the more common ones are about the size and shape of a fountain pen and are provided with a pen clip hence they are sometimes referred to as pen meters or pocket meters The outer container is usually made of plastic and the inside chamber wall and central collecting electrode are made of conducting plastic or of aluminum coated with graphite This inside chamber serves as an air capacitor or ion chamber The central electrode is insulated from the outer cylinder wall by polyethylene or some other suitable insulating material An electric charge can be placed on the central electrode by removing an end cap and applying a voltage of 100–200 volts between the electrode and the chamber wall

Dosimeters, when charged, are essentially air capacitors and the amount of discharge during use is proportional to the absorbed dose of x or γ radiation received The usual dosimeter is relatively energy independent for radiation from 300 kev to 2 Mev but often reads high by a factor of 2 to 3 at about 100 kev Some dosimeters have a thin wall and respond qualitatively to β radiation but the usual wall thickness is such that it excludes β rays of energy less than about 300 kev Plastic walled dosimeters respond to fast and thermal neutrons but give quantitative information only for x and γ radiation

In the common dosimeters, called pocket meters the electrical discharge (and proportionate absorbed dose) is measured by placing the dosimeter in a fiber electrometer and measuring the voltage drop across the chamber In some dosimeters the fiber electrometer is built into one end of the chamber (*see illustration*) and the discharge of the dosimeter can be determined by pointing the window end toward the light and looking through the lens at the image of the fiber projected on a scale marked off in millirads (mrad) Also some models of fiber dosimeters have a built in charging mechanism (usually an electrostatic friction charger) with which the meter may be charged periodically

All dosimeters discharge slowly because of insulator leakage This leakage may range from a few millirads per week for a good meter to 50 mrad/day (or more) for a poor meter Some dosim



Pocket fiber electrometer

eters leak badly during humid weather, and others discharge completely when dropped However, a well designed dosimeter will give reliable results under most operating conditions The useful sensitivity for dosimeters is about 5–300 mrad, however, this sensitivity may be varied over a wide range by modifying the electrical capacitance *See* FILM BADGE; MONITORING (IONIZING RADIATION)

[K Z M]

Doublet flow

In hydrodynamics a doublet is the combination of a source and a sink of equal strength which are allowed to approach each other in such a manner that the product of their strength and the distance between them remains constant in the limit (*see* SOURCE FLOW; SINK FLOW) Doublets have directional properties, the line drawn from the sink toward the source being the axis of the doublet The strength of a doublet is proportional to the product of strength of source and distance between source and sink before the limit is taken

The doublet is a flow element that is used in combination with other elements to build up special flow cases For example, a uniform flow superposed on a two dimensional doublet so that the axis of the doublet is directed upstream yields the flow case of uniform flow around a circular cylinder In three dimensional flow, uniform flow and a doublet directed upstream result in the case of uniform flow around a sphere

The flow net consists of equipotential lines and streamlines such that the change in value between adjacent lines is the same In two dimensional flow, the streamlines are circles with centers on the y axis and the equipotential lines are circles with centers on the x axis (Fig 1) In three dimensional flow, the flow net is somewhat similar, except that the closed loops are not circles (Fig 2) *See* STREAMLINE FLOW

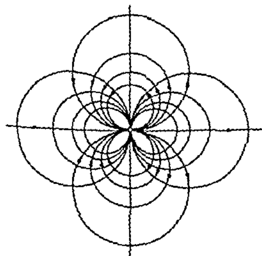


Fig 1 Equipotential lines and streamlines for the two-dimensional doublet

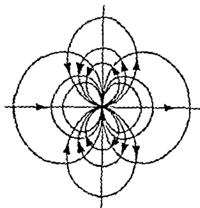


Fig 2 Streamlines and equipotential lines for a three dimensional doublet

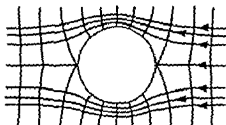


Fig 3 Streamlines and equipotential lines for uniform flow about a sphere at rest

Three dimensional doublets may be distributed along lines over surfaces, or through volumes in such a way that the strength per unit length, area or volume is finite For example a uniform distribution of doublet strength over the periphery of a circle with axes normal to the plane of the circle yields flow around a torus-shaped body when a uniform flow is superposed in the direction of the negative axes of the doublets

Steady flow around a sphere is the result of doublet flow and flow at uniform velocity (Fig 3) The equation for velocity potential ϕ is given by

$$\phi = \frac{Ua^3}{2r^2}\cos\theta + Ur\cos\theta$$

The first term on the right is the velocity potential for a three-dimensional doublet with axis in the $+x$ direction and the second term is for uniform velocity U in the $-x$ direction The spherical polar coordinates are r and θ and a is the radius of sphere

[V L S]

Douglas-fir

A large coniferous tree, *Pseudotsuga menziesii* (formerly *P taxifolia*), also known as red fir This tree may reach a height of 300 ft and is next in size to the giant sequoia and the redwood (*see* REDWOOD SEQUOIA) It grows in the Pacific Coast region and the Rocky Mountains of the United States and differs from true fir in having short leaf stalk elliptical leaf scars and pendent cones This g is characterized by bracts extending out bey cone scales *It is an important timber t*



Douglas fir *Pseudotsuga menziesii* (A H Graves Illustrated Guide to Trees and Shrubs Harper 1956)

ranks first in the United States in total stand and production of lumber and veneer for plywood The wood is hard and strong and is used throughout the United States for construction timber It is also used for cooperage mine timbers mill work rail way car construction flooring furniture ships and ladders In addition it is used as a shade tree an ornamental and for shelter belts Many cultivated varieties are planted in the eastern United States and Europe Douglas fir forests in the United States contain about 400 000 000 000 board ft of lumber More than three fourths of it is in western Washington and western Oregon The annual cut is about 10 000 000 000 board ft and exceeds that of southern pine which formerly was first in lumber production See FOREST AND FORESTRY TREE [A H G]

Dove

Any of over 300 species of birds of the cosmopolitan family Columbidae Some are called pigeons although there is no real difference between doves and pigeons Thus *Columba livia* commonly known as the domestic pigeon is called rock dove by ornithologists Another common species is the band tailed pigeon *C fasciata* (Fig 1) The various domestic pigeons are all races of this bird It is



Fig 1 The band tailed pigeon *Columba fasciata* length to 15½ in (From E L Palmer Fieldbook of Natural History McGraw Hill 1949)

native to Europe Africa and Asia and is now well established in the United States in the feral state often becoming a pest around public buildings

The passenger pigeon *Ectopistes migratorius* was once one of the world s most abundant birds migrating across the United States in flocks of enormous size However a combination of over shooting habitat destruction and other factors wiped it out The last known passenger pigeon died in captivity in 1914



Fig 2 The mourning dove *Zenaidura macroura* length to 11 in (From E L Palmer Fieldbook of Natural History McGraw Hill 1949)

Similar to the passenger pigeon but smaller is the common widely distributed mourning dove *Zenaidura macroura* (Fig 2) It is a game bird of increasing importance in the southern United States and is also gaining acceptance as a game bird in the central states

There are 11 other doves and pigeons in the United States most of them limited to the southern states See COLUMBIFORMES [J D B]

Dracunculoidea

A group of parasitic nematodes characterized by their morphology by their habitat in the tissues of the host and by the way the larvae which are produced by viviparous females leave the host through a lesion in the skin This group is considered to have the taxonomic status of an order or superfamily according to specialists who study this group

Dracunculus medinensis This species is called the Guinea worm or fiery serpent and has been known for ages as a human parasite in Asia and Africa The female worm often a yard in length lies in the tissues under the skin When the worm is mature a blister forms over its head usually on the foot or lower leg of the host In a few days the blister bursts and when the lesion is immersed in water many larvae pour out This outpouring of larvae may be repeated many times until the worm is empty If appropriate species of water fleas *Cyclops* are present in the water they swallow the larvae which then undergo development to the in

fective stage in a few weeks Humans become in fected by swallowing the infected Cyclops in drink ing water Experimental infections in dogs have shown that the larvae are freed by digestion of the Cyclops The larvae penetrate the wall of the in testine into the loose connective tissue and later m grate by the lymphatics to the skin The devel opment of the female worms requires nearly a year The males are much smaller and few in num ber It is not known whether they fertilize the females or if the larvae develop parthenogenetically In humans when the female is localizing under the skin severe allergic symptoms are produced until the blister breaks

In India high infection rates result from the use of step well into which people wade to obtain household water Rebuilding of wells and chemical treatment of ponds are suggested for control Mod ern treatment has been devised but the ancient picturesque but gruesome method of slowly pull ing the worm out by rolling it on a stick is still in use

Other species of *Dracunculus* have been found in mammals and reptiles in various countries Species of *Micropleura* occur in crocod les Species of *Ph lometra* are found in fishes and are also transmitted by *Cyclops* *See* NEMATODA [J A.S]

Bibliography C F Craig and E C Faust *Clini cal Parasitology* 5th ed 1951 V N Moorthy *Am J Hygiene* 27 437 1938

Drafting

The making of drawings of objects structures or systems that have been visualized by engineers sci entists, or others Such drawings are usually made with mechanical drawing instruments but may be sketched freehand

Drafting is done by persons with varied back grounds Engineers often draft their own des gns to determine whether they are workable structur ally sound and economical On the other hand much routine drafting is done under the supervi sion of engineers by technicians specifically trained as draftsmen *See* DESCRIPTIVE GEOMETRY ENGI NEERING DRAWING [C J B]

Drafting machine

A movable straight edge that parallels a fixed horizontal base line In making mechanical line drawings on a drawing board an easily movable horizontal straight-edge is helpful One solution to the problem is the univer al drafting machine The one shown schematically employs two four bar parallel linkages with one l nk common to the two linkages If the mounting bracket link 1 is horizontal then link 3 must remain horizontal as indicated by the dotted line Links 3 and 5 are in tegral being formed by the elbow plate Link 5 is always vertical and link 8 therefore must remain vertical Horizontal and vertical straightedges usu ally pivoted to reduce the need for triangles are attached to link 8 A popular variation of this linkage is one in which elements of flat rim wheels of equal diameter are kept parallel by metal bands around each pair of wheels [E S F]

1
2
4
elbow plate
6
8
5
3
7

Un versal drafting mach ne

Dragonfly

Any member of the suborder Ani optera order Odonata Dragonfl es are among the best known insects and are especially common in the southern part of the United States where they are frequently called snake doctors snake feeders and devil s darning needles The first two names are in recognition of the widely held superstition that dragonflies tend injured and ailing snakes Dragon flies have two pairs of large net veined wings held at right angles to the slender body when at rest The related damselflies of the suborder Zygoptera are similar but more fragile and fold their w ngs over the back when at rest

The ten spot dragonfly *L bellula* sp w ngspread 1½ n (From E L Palmer *F eldbook of Natural H story* McGraw H ll 1949)

Dragonfly and damselfly nymphs are aquatic and active predators eating insects tadpoles, and young fish Adults eat insects which they catch the w ng The metamorpho is is t Odonota are world wide in distribution w it 4500 known species. They are most comr tropical and subtropical regions. *See*

Drawing of metal

A stretching operation the through a d e resul dimensions (Cupping drawing are similar formed on heet metal other article *See* SHEET

Rod wire and tube drawing are cold working processes used primarily to reduce the cross section of the stock improve surface finish and dimensional tolerances and strengthen the metal by strain hardening

Rod and wire drawing For drawing stock over ¼ in in diameter a draw bench is used This consists of a frame 50–100 ft long with a die holder at one end a driving mechanism at the other end and a chain driven carriage (equipped with jaws to hold the rod) which moves on tracks on the frame

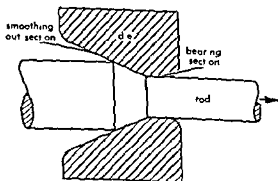
For wire sizes over ¼ in in diameter single block and holder units called bull blocks are used The block is a large power driven drum 6–32 in in diameter The die is held a few inches from the block so that the center line is tangent to the block

For sizes under ¼ in in diameter a wire drawing frame containing several blocks mounted with driving and control mechanisms and die holders is used The wire is coiled around each block several times before it passes to the next die

Before the rod is drawn through the die it is surface treated by pickling to remove scale Steel rod is then coated with lime copper or tin Lime is used with greasy or soapy lubricants the metallic coatings are used with water or water plus flour and rye meal Less reduction per draft is possible in wet drawing however a greater total reduction without intermediate annealing is possible For drawing copper wire no special coating is applied the wire is treated to remove scale and a greasy lubricant is applied at the die

The drawing die is usually made of tungsten carbide (diamond is used for fine wire) inserted in a steel die block The entrance portion of the die hole is not in contact with the rod but is filled with lubricant A tapered portion of the die called the smoothing out section is the region where plastic deformation occurs (as illustrated) This section must be properly tapered and very smooth A short bearing section (either cylindrical or slightly tapered) helps maintain dimensions as the die wears

As the stock is pulled through the die work is required to obtain uniform elongation overcome friction and produce nonuniform shear deformation in the outer layers The work required to produce shear deformation decreases with decreasing die angle and with increasing reduction. Friction, however increases with decreasing die angle Thus an optimum die angle exists which varies with the drawing conditions and the particular metal being drawn.



Cross section of drawing die

Because the plastic flow is nonuniform across the rod residual stresses are always present after the drawing is completed The residual stresses at the surface are generally tensile in the case of heavy reductions and compressive for light reductions the magnitudes sometimes exceeding one half of the yield strength These stresses may be relieved by stretching roller straightening or annealing

Tube drawing The outside diameter of a tube is reduced and the wall thickness is generally increased slightly by pulling the tube through a die without the use of an internal mandrel (called sinking) By using a stationary mandrel (plug) the wall thickness can be reduced along with the reduction in diameter In some instances a floating plug or a long cylindrical rod may be used as a mandrel The cross sectional shape of the tubing may be changed by use of appropriate dies and matching mandrels *See* METAL FORMING.

[R L F]

Bibliography J M Camp and C B Francis, *The Making Shaping and Treatment of Steel* 6th ed 1951 G Sachs and K R Van Horn *Practical Metallurgy*, 1940

Dredge

A floating excavator used for widening or deepening channels building canals constructing levees or raising material from stream or harbor bottoms to be used elsewhere as fill Dredges are of either hydraulic or mechanical types

In the hydraulic unit a large suction pipe supported and moved about by a boom is lowered to the bottom A mechanical agitator or cutter head churns up the earth immediately in front of the suction pipe and centrifugal pumps mounted on the dredge suck up both water and loose solids The material is sometimes discharged into the hold of the dredge itself or is placed into barges which can be towed to disposal areas In most cases however the dredged material is pumped directly into floating pipelines and carried to areas outside the channel or to land

The most common mechanical unit is a dipper dredge which works much like a land power shovel A hinged bottom bucket at the end of a dipper stick scoops material from the bottom and loads it into dump bottom barges for subsequent disposal Another common mechanical unit is the bucket dredge It carries numerous buckets mounted on an endless chain and works much like a trencher or ditching machine Other mechanical dredges use clamshell or dragline buckets and some European models use a large rotating wheel on which are mounted a number of buckets

[E M Y]

Drier (paint)

A material which facilitates the oxidation of oils Driers are salts of metals most commonly lead manganese or cobalt The acids used to make the salts may be drying oil fatty acids rosin or naphthenic or octoic acids from petroleum The only function of the acid portion is to render the metal soluble in the oil

A drier will cause linseed oil which in a pure state requires about 3 days to dry to a hard film to dry in about 4 hours The mechanism of drier reaction is not entirely clear but it is believed that the formation of peroxides during the oxidation of the oil and their subsequent destruction to form radicals suitable for polymerization are essential parts of the drying mechanism and that catalytic amounts of drier metal facilitate the destruction of the peroxides and the formation of free radicals However there is evidence that not all drier metals act through the same mechanism because mixtures of driers will produce faster drying than can be obtained with any amount of a single drier

Cobalt is the most reactive of the drier metals and is used in amounts of less than 0 1% of metal based on the oil It is regarded as a surface drier and is widely used for thin films When thicker films are encountered one of the other metals is often used as a through drier to dry the bulk of the film whereas cobalt is used to dry the surface and *free it from tack*

Lead is much less reactive than cobalt and is used in amounts of about 0 5% of the oil Lead is rarely used alone but with cobalt or manganese as the through drier portion of the drier

Manganese is somewhat similar to cobalt but less reactive Because it will discolor certain oils it is more commonly used in exterior paints where air and sunlight will bleach it

Iron is not an effective drier at low temperatures but is often used in baking finishes Iron salts carry enough color that they cannot normally be used in white paints

Zinc and calcium do not dry oil films by themselves but are often used as auxiliary driers with one of the other metals These salts are also effective as wetting agents

Numerous other metals including cerium vanadium and zirconium have been used occasionally and are effective driers but their use is not economically justifiable except for a number of special cases

Certain organic compounds such as 1 10-phenanthroline also catalyze the drying of oils and have been used for this purpose The mechanism by which they operate is not clear but they have a definite place especially where freedom from metallic contaminants such as lead is required

Originally driers were made by cooking salts or oxides of the appropriate metals with fatty acids oils or rosin Oils which had been cooked with metals were called boiled oils, because the reaction releases water resulting in a boiling of the

Later means of producing the appropriate organic salts by reaction in solution were developed and driers are normally sold as solutions in mineral spirits containing a definite amount of metal so that additions can be made accurately As sold to the paint industry driers are usually solutions of the salts of a single metal but painter s drier sold to users to add to paints is a mixture of several metals chosen to produce the effect desired

Although small amounts of drier are essential to the formation of a satisfactory paint film in a reasonable time larger amounts are apt to lead to premature embrittlement and failure of the paint film and should be avoided *See* DRYING OIL PAINT

[F SD]

Drilling machine

A motor driven device fitted with an end cutting tool that is rotated with sufficient power either to create a hole or to enlarge an existing hole in solid material One or more flutes or grooves in the drill tool conduct coolant to the cutting lips and also provide chip relief

Twist drills with two spiral flutes are commonly used to originate holes while 3 and 4 flute non centercutting drills are used to enlarge holes Other types are center core hog nose and gun barrel drills Cylindrical saws or pin drills are used in trepanning to cut large circular holes

Drilling machines range in size and complexity *from small sensitive* drill presses through upright and multiple spindle models designed for mass production as illustrated Large radial types handle special jobs as illustrated

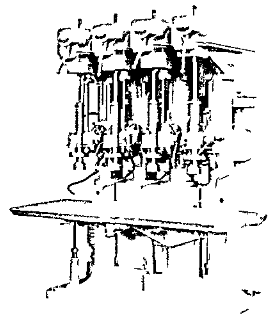


Fig 1 Four spindle drilling machine (*Fosdick Machine Tool Co*)

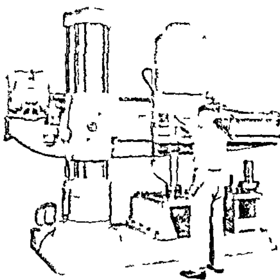


Fig 2 Rad al drilling mach ne set up for trepanning (Cincinnat Bickford)

Drilling speeds usually decrease with material hardness while feed per revolution increases with drill diameter Spotfacing to finish the area around a hole counterboring to enlarge the diameter over part of the depth and countersinking to chamfer edges of a hole are operations frequently performed during drilling set ups See BORING MACHINING OPERATIONS REAMER [A P]

Drone

A pilotless aircraft subordinated to the controlling influences of a remotely located command station Early guided weapons including numerous types under development in the closing stages of World War II were essentially droned vehicles adapted to a destructive mission With the extension of missile armament the term guided missile was universally *adopted* by the layman to describe any remotely directed or internally guided airborne instrument from continuously controlled drones to ballistic missiles Specifically a drone is a *continuously and* remotely controlled pilotless aircraft *capable of* performing any nondestructive mission when used destructively the device *is properly* termed a missile

Evolution Rudimentary principles of drone operation were applied in vehicles *developed by* E Sperry C Kettering A Low G de Havilland and H Folland during World War I and *essential* elements of remote controllability *were* success fully introduced during the 1920s *and 1930s*

Developmental work in *America* was seriously restricted until 1935 when the *experimentation* of amateur model airplane builders *produced* practical equipment in the field of *radio*-wave guidance

In Great Britain a *logical progression from* prior developments at the Royal Aeronautical Establishment produced first the A W *Wolf* and then the Larynx of 1927–1930 B th *responded to* radio control the latter embodied gyro stabilization

In 1934 the Queen aircraft of the Royal Air Force included the Fairey IIIF conversion de Havilland s Queen Bee and the Airspeed Queen All were launched by catapult and recovered by guiding the planes to a normal landing usually on the water

By 1940 the Radioplane Company in California had developed a quarter scale radio controlled airplane for the military services to function as a gunnery training device World War II production of similar target drones by various manufacturers amounted to thousands

Principles of operation Droned aircraft normally employ one of two basic principles of *remote* guidance either radio control or radar control Essentials of any guidance system include the ability to control the vehicle's attitude and to establish a flight path For simple missions flown within sight of the controller drones are equipped with a receiver and signal converter or amplifier to operate control servo devices Many drones like *that shown* in Fig 1 fly at altitudes and slant ranges *well be*yond the sight of the controller Gyro devices and characteristics inherent in the design of the *craft* combine to stabilize the drone in pitch roll and yaw attitudes Command signal inputs may override *the effectiveness of these sensors* to impart flight path deviations Coupled with this *is the* importance of maintaining a visual plot of the drone's range azimuth and altitude, this is *generally* accomplished through radar tracking *and a graphic* presentation

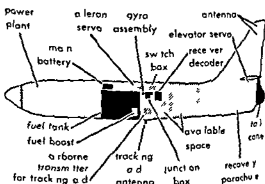
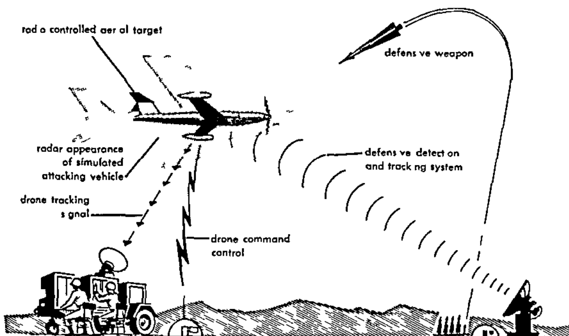


Fig 1 Inboard prof le of typical drone *illustrat ng* internal d sposit on of equipment

Propulsion systems enjoy virtually *unlimited* applications in drones inasmuch as *performance* requirements are extremely flexible *Piston engines* turboprops turbojets, ramjets turboramjets pulse jets solid rockets and liquid rockets have *all been* employed with varying degrees of success in drones Low performance types are *generally* powered with small reciprocating engines For missions in the medium performance class small turboprop pulsejet solid rocket or liquid rocket *propulsion* systems provide the motive force *To attain the high* performance category prevailing *by 1960* powerful turbojets rockets or ramjets *became a necessity*



Fg 2 Typ cal drone m ss on situat on depict ng command control track ng detect on and intercept on

Applications Drones have gained their widest acceptance as targets for antiaircraft gunnery and missile development evaluation and training (Fig 2) Obsolete military aircraft have been converted to drone operation for these purposes and for cloud sampling or blast effects in conjunction with nu clear device tests As an adjunct to successful at tack by air launched ballistic missile weapons the diversionary or decoy drone may be indispensable Few vehicles can approach the versatility of drones for reconnaissance and surveillance missions

[S E W]

E [illegible]
and [illegible]
For [illegible]
Principles of Guided Missile Design 1955 A R Weyl *Guided Missiles* 1949

Drought

A deficiency of rainfall seriously hampering growth of farm crops or other vegetation Such deficiency may be local as over a few counties or wide spread like the great droughts of 1894–1895 and 1934 that affected most of the United States

Drought is indicated but not uniquely defined in terms of rainfall deficiency in inches percentage of normal and number of consecutive days without rain But the types of vegetation the season and the character and condition of soil are other de- termining factors Also windiness low humidity sunshine and high temperatures contribute to dry ness

The term drought is sometimes applied to a pro- longed deficiency of water for domestic or irriga tion needs These shortages may be largely in streams but may include low underground water supply

In very wet regions an extended period with less than normal rainfall may not be a drought if nor mal needs are met

Meteorological conditions for droughts on a large scale involve atmospheric circulation over at least a hemisphere and possibly the world Air flow gen erally determines the amount of moisture trans ported into a region and also where cyclones fronts, or thermal instability will release the moisture as precipitation A prolonged absence of northward air flow at low levels from the Gulf of Mexico will produce drought over much of the United States east of the Rockies Even with ample air flow there may be local drought in some areas and floods in others because of the locations or paths of cold air and cyclones then prevailing

I R Tannehill showed that rainfall over the United States is generally less than normal in years when atmospheric pressure on the Pacific coast is relatively high He suggested that abnormally cold water off the coast in those years may be a factor This drought effect extends over most of the United States but not eastern Canada presumably because storm tracks are farther north than normal *See* PRECIPITATION (METEOROLOGY) [J R F]

Bibliography I R Tannehill *Drought* 1947

Drug resistance

A decreased reactivity of living organisms to the injurious actions of certain chemicals Resistance is usually specific for one chemical or for a re lated group of chemicals while the susceptibilities to others remain unchanged Independent resist

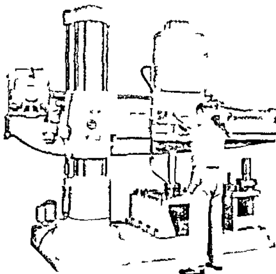


Fig 2 Radial drilling machine set up for trepanning (Cincinnati Bickford)

Drilling speeds usually decrease with material hardness while feed per revolution increases with drill diameter Spotfacing to finish the area around a hole counterboring to enlarge the diameter over part of the depth and countersinking to chamfer edges of a hole are operations frequently performed during drilling set ups *See* BORING MACHINING OPERATIONS REAMER [A T]

Drone

A n l lesa f l

under development n the l
W :
a
ar
ac
rected or internally guided airborne instrument from continuously controlled *drones* to ballistic missiles Specifically a drone is a continuously and remotely controlled pilotless aircraft capable of performing any nondestructive mission when used destructively the device is properly termed a missile

Evolution Rudimentary principles of drone operation were applied in vehicles developed by F Sperry C Kettering A Low G de Havilland and H Folland during World War I and essential elements of remote controllability were successfully introduced during the 1920s and 1930s

Developmental work in America was seriously restricted until 1935 when the expe
ama
cal
l a logical progression from prior developments at the Royal Aeronautical Establishment produced first the A W Wolf and then the Larnyx of 1927-1930 Both responded to radio control the latter embodied gyro stabilization

In 1934 the Queen aircraft of the Royal Air Force included the Fairey IIIF conversion de Havilland s Queen Bee, and the Airspeed Queen All were launched by catapult and recovered by guiding the planes to a normal landing usually on the water

By 1940 the Radioplane Company in California had developed a quarter scale, radio controlled airplane for the military services to function as a gunnery training device World War II production of similar target drones by various manufacturers amounted to thousands

Principles of operation Droned aircraft normally employ one of two basic principles of remote guidance either radio control or radar control Essentials of any guidance system include the ability to control the vehicle's attitude and to establish a flight path For simple missions flown within sight of the controller drones are equipped with a receiver and signal converter or amplifier to operate control servo devices Many drones like that shown in Fig 1 fly at altitudes and slant ranges well beyond the sight of the controller Gyro devices and characteristics inherent in the design of the craft combine to stabilize the drone in pitch roll and yaw attitudes Command signal inputs may override the effectiveness of these sensors to impart flight path deviations Coupled with this is the im

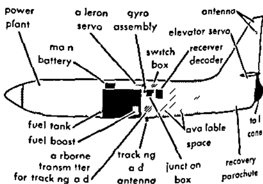


Fig 1 Inboard profile of typical drone illustrating internal disposition of equipment

portance of maintaining a visual plot of the drone's range azimuth and altitude, this is generally accomplished through radar tracking and a graphic presentation

Propulsion systems enjoy virtually unlimited applications in drones inasmuch as performance requirements are extremely flexible Piston engines turboprops turbojets ramjets turboramjets pulse jets solid rockets and liquid rockets have all been employed with varying degrees of success in drones Low performance types are generally powered with small reciprocating engines For missions in the medium performance class small turboprop pulsejet solid rocket, or liquid rocket propulsion systems provide the motive force To attain the high performance category prevailing by 1960 powerful turbojets rockets or ramjets became a necessity

Zinc electrode Zinc used in dry cells should be free from impurities except for small amounts of lead and cadmium which improve the mechanical properties

Large zinc cans are made by forming a cylinder from rolled sheet zinc and soldering a lapped seam To this is also soldered a bottom *circular* disk Smaller sizes are made by drawing *or* by impact extrusion The weight of zinc used in several sizes of zinc *cans ranges* from 110 grams *(g)* for the no *6 to* 5 g for the AA size These weights corre spond to a theoretical capacity of *90 amp hr* and 4 1 amp hr respectively Actually the cell capacity is much *smaller*

Black mix The black mix is composed of man ganese dioxide mixed with carbon *black* The manganese dioxide is usually obtained from natu ral ore (African) but may be a synthetic prod ict prepared by chemical precipitation or by electro chemical *methods* Mixtures *of the natural* and synthetic oxides are also used The carbon black is usually acetylene black made *by the thermal* de composition of acetylene Graphite is used to a lesser extent Manganese dioxide has a *theoretical* capacity of about 0 3 amp hr/g The practical capacity is somewhat less The *carbon* black is used in varying proportions depending *on* design factors between one fifth and one *tenth* of the weight of the manganese dioxide *In* addition to these components the black mix also contains electrolyte amounting to about 25% of the total *weight*

Carbon electrode The carbon rod used *in a* cy lindrical cell serves as the *conductor of* electricity for the positive electrode it also serves as a vent to allow gas to escape *Carbon* rods are usually made from petroleum coke which is calcined ground and mixed with pitch *The* green rods are baked to form a hard carbon *having* low elec trical resistance They may be *partially* water proofed to prevent capillary creepage *of* electro lyte out of the cell

Flat cells are usually made *with* duplex elec trodes The zinc is coated on one *side* with a car bonaceous coating which *serves* to conduct elec tricity between the zinc and the black mix of the adjacent cell

Gelatinous paste The paste is commonly made from a mixture of electrolyte with corn starch and wheat flour More *recently* developed synthetic materials have given excellent results Methyl cel lulose *has been* reported to provide lower electric resistance and better keeping quality

Starch pastes are added to the *cell in* liquid form One type of paste is gelatinized by heating the cell in a water bath *or in a current of* steam Cold setting pastes gelatinize at room tempera ture

Dry cells are also made with the paste *applied* to a bibulous paper The paper *after treatment* is wrapped around the compacted black mix called the bobbin or core The wrapped bobbins are placed in the zinc cans and *sometimes more* elec trolyte is added subsequently *Flat* cells are made by *placing treated* paper *containing* the paste be tween the black mix cake and the zinc of each cell

Electrolyte The electrolyte is made by dis solving ammonium chloride and zinc chloride in water A very small *amount* of mercuric chloride is usually added This *component* however con verts to zinc chloride *as* soon as the zinc and electrolyte come into contact *Mercury* then plates out on the zinc *The* composition of the electrolyte depends on the cell designs

During discharge the *composition* of the elec trolyte changes *In one* test in which a D size cell was discharged *through* a 4-ohm resistance the pH of the paste layer next *to the zinc* changed from 5 7 to 3 8 (more acid) while the pH of the *innermost portion of* the mix went from 5 8 to 10 1 (more alkaline)

Ordinary *dry cell* electrolyte has a resistivity of 2 42 ohm cm at +20°C For low temperature oper ation special electrolytes *have been* developed *An* electrolyte of 12% zinc chloride 15% lithium chloride 8% *ammonium* chloride and 65% water is fluid at −40°C Other electrolytes for low *temperature operation* use a mixture of *calcium* chloride zinc chloride and ammonium chloride solutions

Cell enclosure While round cells may use the zinc can *as the* enclosure this *zinc can may be* placed in a paper tube and then within an outer *jacket of sheet* steel The top of the zinc *can is* closed with a washer and asphalt sealing com pound The steel *jackets are formed to the* re quired dimensions and the cells are finally sealed in these *containers* with a metal top closure *No* zinc is exposed Plastic coated paper jackets are also used *as the outer* enclosure

Cells using the zinc can as the outer enclosure are *provided with* a top closure An *impregnated* cardboard washer is slipped over the central car bon *rod to center* it and to support the top seal Wax or pitch seals are poured while hot to fill the space *between* the top of the washer and *the top* of the cell

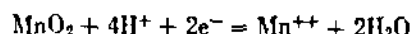
Flat cells use thin plastic wrappings around the [illegible] cell This confines the electrolyte to [illegible] a [illegible] le [illegible] ll [illegible] he stack is bound together by [illegible] in molten wax for further *moistureproofing*

Cell chemistry At the anode (zinc) the zinc oxidizes to *zinc ion* and simultaneously gives off electrons to the external circuit at a rate pro portional *to the current* For each ampere *which* flows 1 2 g of zinc per hour are converted to zinc ion

At the cathode (manganese dioxide) the elec trons *from the* external circuit reduce the ma ganese *dioxide* to three different substance

pending on circumstances which have not yet been thoroughly explained Studies have shown, however that the total ampere hour output of the cell can be accounted for by analyzing the cathode mix for the following substances soluble manganese (Mn^{++}) each gram of which accounts for nearly 1 amp hr of discharge insoluble manganite ($MnOOH$) each gram of which accounts for about 0 3 amp hr of discharge insoluble hetaerolite ($ZnO \cdot Mn_2O_3$) each gram of which accounts for about 0 22 amp hr of discharge

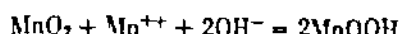
The electrochemical reduction of the manganese dioxide (MnO_2) has been reported to occur as the reaction to form soluble manganese



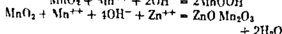
$$MnO_2 + 4H^+ + 2e^- = Mn^{++} + 2H_2O$$

This occurs only when the cell delivers *current*

Two secondary reactions can then occur



$$MnO_2 + Mn^{++} + 2OH^- = 2MnOOH$$



$$MnO_2 + Mn^{++} + 4OH^- + Zn^{++} = ZnO \cdot Mn_2O_3 + 2H_2O$$

This last reaction can occur only if zinc is in solution in the cathode mix *See* ELECTROCHEMISTRY

Dry cell operating characteristic Table 1 shows the approximate characteristics of commercial D size dry cells on continuous discharge to a final voltage of 1 volt per cell at 70°F

Table 1

Hours to 1 volt	Amp	Amp-hr	Amp-hr/lb	Relative output %
1	0 48	0 48	2 27	20
10	0 11	1 10	5 20	46
100	0 024	2 40	11 36	100
1000	0 0045	4 50	21 3	187

The cell voltage on discharge decreases continuously The available capacity *therefore* depends on the choice of end voltage *This is* particularly important for high discharge rates The output of a D size cell *on a 4-ohm* load for different final voltages is shown in Table 2

Table 2

Final voltage	Time hours	Amp-hr	Watt hrs	Watt hrs/lb
1 4	0	0	0	0
1 2	0 75	0 244	0 317	1 5
1 0	2 35	0 681	0 792	3 75
0 9	3 50	0 955	1 04	4 93
0 8	5 50	1 37	1 37	6 50

Special batteries have been built *with a* much smaller change in output with current *These* batteries are not available commercially but are cited to show the possible high-drain *output of* Leclanche cells By greatly increasing the electrode area with respect to the mix thickness the output changes much more slowly as *shown* by comparison of Table 3 with Table 1

Table 3

Hours to 1 volt	1	10	100
Amp hr	0 029	0 045	0 0 0
Ratio to 100 hour output	0 42	0 64	1 00

On intermittent discharge, the total output is usually greater than on continuous output Thus a D size cell on 5 min discharge per day through 4 ohms will deliver about 3 4 amp-hr to 0 75 volt On continuous discharge, the cell delivers about 1 6 amp hr to 0 75 volt

Temperature has a pronounced effect on dry cell output In general the performance is better as the temperature increases As the temperature decreases the cell potential measured by an electrometer drops about 0 0004 volt/°C The working voltage however, drops much more The greater the drain rate the greater is the effect of temperature For an F size cell the effect of temperature on initial discharge voltage for various drain rates is shown in Table 4 The change in voltage per degree centigrade is roughly proportional to the current for a given cell

Table 4

Current	Voltage at +30°C	Voltage at −30°C	Change in voltage per C°
0	1 645	1 610	0 000 8
0 020	1 641	1 490	0 0025
0 050	1 632	1 280	0 00[illegible]9
0 075	1 621	1 10	0 0087
0 100	1 612	0 96	0 011
0 140	1 605	0 86	0 012

The effect of temperature on flash current is shown in Table 5 Flash current is the maximum current delivered to a load of 0 01 ohm

Table 5

Temperature °C	25	0	−30
Relative flash current %	100	66	0

The effect of temperature on hours of continuous service is shown in Table 6 A cell which delivers 500 hours at 70°F will deliver 310 hours at 40°F 135 hours at 0°F, 55 hours at −15°F If it delivers only 20 hours at 70°F, it will deliver 8 hours at 40°F, etc

Table 6

Temperature	Hours	Hours
70°F	500	20
40°F	310	8
0°F	135	3
−15°F	55	0

Better low temperature output can be obtained with special electrolytes and cell structures giving a high ratio of electrode area to mix thickness and special types of manganese dioxide

Shelf life The *preceding* characteristics are those of fresh cells not over 3 months old Deterioration takes place in an idle cell so that the o t

put decreases as the cell ages The limiting storage time after which cell performance is not likely to be satisfactory is called the shelf life of the cell

Deterioration in a dry cell occurs in a number of ways (1) Zinc can oxidize by reaction with the electrolyte this reaction produces hydrogen (2) Manganese dioxide can be reduced by carbon and by the organic materials used in the cells this can produce carbon dioxide (3) Water can be evaporated from the electrolyte this increases the cell resistance and alters the composition of the electrolyte unfavorably

Table 7 gives the results of storage at various temperatures, as shown by extensive tests at the Naval Ordnance Laboratories Figures in the table are per cent capacity loss per year Table 7 shows that the per cent loss per year diminishes as the storage temperature decreases It is less at the very low drain (4-month rate) than at the high drain (1 hour rate) It is less for round cells than for flat cells

Table 7

Storage temperature °F	Discharge rate (to 1 1 volt)	% loss per year	
		Flat cell	Round cell
70	1 hour	41	28
	4 month	32	13
40	1 hour	26	17
	4 month	14	5 1
10	1 hour	19	18
	4 month	11	2 9
−30	1 hour	6 8	2 7
	4 month	6 6	

Another factor is the end voltage to which the discharge is run Discharging to 0 9 volt showed about 70% of the loss obtained by discharging to 1 1 volt [S EI]

Bibliography *Specification for Dry Cells and Batteries* Natl Bur Standards Circ 559 1955 M E Wilke *Trans Electrochem Soc* 90 433 1946

Dry ice

A solid form of carbon dioxide CO_2 which finds its largest application as a cooling agent in the transportation of perishables It is nontoxic non corrosive sublimes directly from a solid to a gas and leaves no residue At atmospheric pressure it sublimes at −109 6°F absorbing its latent heat of 246 4 Btu per pound Including sensible heat absorption the cooling effect per pound of dry ice is approximately 270 Btu at storage temperatures above 15°F and 250 Btu at lower temperatures Slabs of dry ice can easily be cut and used in ship ping containers for frozen foods for cooling refrigerated trucks and as a supplemental cooling agent in refrigerator cars *See* CARBON DIOXIDE

The manufacture of carbon dioxide gas is a chemical process The gas is liquefied by compressing it to 900–1000 pounds per square inch gage in three stages of reciprocating compressors and then condensing it in water-cooled condensers The liquid is expanded to atmospheric pressure where its temperature is below the triple point (−69 9°F) The result is the sublimation of carbon dioxide snow which is very porous The snow is removed from the expansion chamber and mechanically compressed into standard 50-lb blocks which measure 10 by 10 by 10 in *See* REFRIGERATED TRUCK REFRIGERATION REFRIGERATOR CAR

[H M HE]

Drydocking

A technique used to remove a ship from the water so that the underwater portion may be inspected repaired maintained or altered Occasionally underwater repairs may be undertaken while a ship is afloat however at regular intervals or as dictated by emergency it may be necessary to expose all of the underwater portion regardless of whether the ship is a small harbor tug or a large transoceanic liner There are three methods of drydocking a ship namely by the use of (1) a marine railway (2) a floating dry dock or (3) a graving dock

Reasons for drydocking The chief reasons for drydocking a ship are to remove marine growths that cause fouling to prevent hull corrosion and to make repairs and alterations

Fouling of the ship s bottom This is caused by marine growths—barnacles mussels and other animal organisms and marine grasses—which attach themselves to the hull Ships plying in warm salt water tend to foul more rapidly than those which ply in colder areas This fouling seriously retards the speed of the ship which must develop additional horsepower to maintain the same speed thus increasing the cost of operation

Prevention of corrosion Corrosion of the underwater hulls of steel ships is caused by electrochemical reactions The most economical way to prevent corrosion is to place a barrier between the steel and the sea water which acts as the electrolyte Paint or other water excluding materials applied directly to the hull will protect the steel as long as the coating remains intact

Galvanic corrosion results from the close proximity of dissimilar metals as is often the case for example in hull piping systems This type of corrosion may be minimized by the use of cathodic protection in which the ship s hull is the cathode and metallic anodes are placed in the area to be protected Such underwater protection is installed replaced and maintained while the ship is dry docked *See* CORROSION MARINE MACHINERY

Repairs and alterations Both periodic and casualty repairs to the underwater hull of the vessel its propellers shafting rudders and sea connections are nearly always made while the ship is removed from the water Alterations to underwater portions of a ship are sometimes advisable to obtain more speed to use less horsepower to obtain the same speed or to improve maneuverability

Frequency of drydocking This is governed by the following factors (1) preservation of the underwater body (2) accidents necessitating repairs to the underwater portion (3) regulations of the U S Coast Guard for American ships (4) regulations of the classification societies such as The American Bureau of Shipping Lloyd s Register of Shipping and Bureau Veritas for ships inspected and classed by them and (5) the trade in which the ship operates and the water she plies Most owners drydock their ships annually for inspection cleaning painting and routine repairs Some owners find it economical to do this semiannually

Types of dry docks The three types of dry docks are known as marine railways floating dry

or steel with rollers on which the ship may be hauled out of the water along a fixed inclined track leading up the bank of a waterway (Fig 1) The advantages of a marine railway lie in the economy of the original construction and the relative low cost of maintenance A marine railway is ideal for ships up to 5000 tons

Floating dry dock This type may be constructed of wood steel or concrete Larger floating dry docks are usually built in sections called sectional floating docks which are secured together by hinge like joints (Fig 2) The dock is submerged to provide the required depth of water over the keel blocks by partially filling its tanks with water The ship to be drydocked is then positioned within the tanks of the dock are rapidly pumped out by powerful pumps located within the dock walls, and the ship is lifted out of the water The water within the well of the floating dock spills out of the open ends of the dock The sectional floating docks may

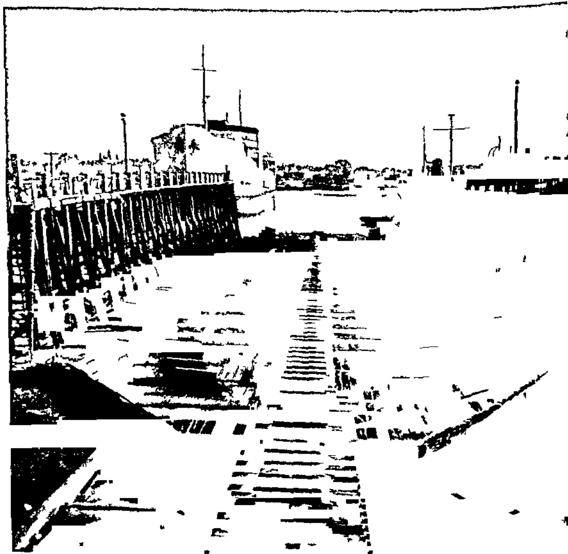


Fig 1 Bow-end view of a marine railway which has been pulled out of the water The tracks on which the rollers travel may be seen at right and left in the foreground There are also center tracks to take the load of the keel

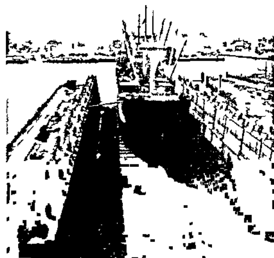


Fig 2 A sectional floating dry dock afloat with a ship in position



Fig 3 A rigid floating dry dock afloat in a graving dock This type of dry dock is unlike the sectional dock in that it is a single structure and must make use of other facilities when underwater maintenance and repairs on the dock itself become necessary

be drydocked for self repair by docking individual sections in the remaining portions

Smaller floating dry docks capable of lifting up to 4000 tons are constructed in one piece and are called rigid floating dry docks (Fig 3)

The advantages of the floating dry dock are low initial cost and mobility This mobility permits its use in any location which has sufficient depth of water to allow for the submerged depth of the dock plus the height of the keel blocks plus the draft of the ship to be drydocked Disadvantages are cost of upkeep need for frequent repairs and the repeated dredgings of the basin necessary to permit its full submersion

Graving dock This consists of an excavation in the ground with a thick concrete base supported if necessary by piling and surrounded on three sides by earth held back by timbers stone cement or steel supports or a combination of these materials (Fig 4) The entrance or seaward end of the dock is usually closed by a caisson of the pontoon type which when flooded is trimmed down into position The caisson is gasketed at the surface of contact and held in position by the pressure of the outside water as the water inside the dock is removed In maneuvering the caisson use is made of pumps located inside on a flat placed above the level reached by the ballast water The dock is flooded by means of valves that extend through the gate or through pipes embedded in the dock walls The caisson its two sides now subject to equal pressure is floated by pumping out the ballast water It is next hauled out of the ship s path by lines led to docks de based capstans The ship enters the dry dock and is positioned over the keel blocks The caisson is then replaced and submerged the

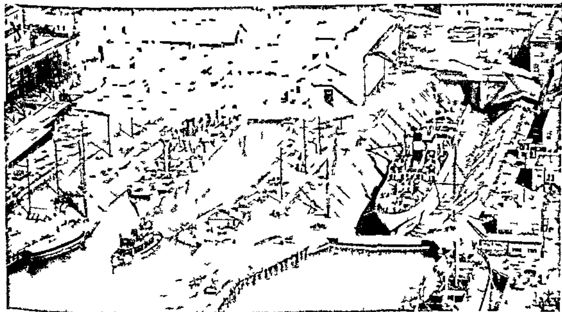


Fig 4 A dry and a flooded graving dry dock The graving dock at the right is dry and work is proceeding on a ship supported by the keel and bilge blocks The graving dock on the left is flooded and the ship is being towed from it This dry dock s gate caisson has been moved to one side

Frequency of drydocking This is governed by the following factors (1) preservation of the underwater body (2) accidents necessitating repairs to the underwater portion (3) regulations of the U S Coast Guard for American ships, (4) regulations of the classification societies such as The American Bureau of Shipping Lloyd s Register of Shipping and Bureau Veritas for ships inspected and classed by them and (5) the trade in which the ship operates and the water she plies Most owners drydock their ships annually for inspection cleaning painting and routine repairs Some owners find it economical to do this semiannually

Types of dry docks The three types of dry docks are known as marine railways floating dry docks, and graving docks The size of the ship usually determines which type is used

Marine railway This consists of a cradle of wood or steel with rollers on which the ship may be hauled out of the water along a fixed inclined track leading up the bank of a waterway (Fig. 1) The advantages of a marine railway lie in the economy of the original construction and the relative low cost of maintenance A marine railway is ideal for ships up to 5000 tons

Floating dry dock This type may be constructed of wood steel or concrete Larger floating dry docks are usually built in sections called sectional floating docks which are secured together by hinge like joints (Fig 2) The dock is submerged to provide the required depth of water over the keel blocks by partially filling its tanks with water The ship to be drydocked is then positioned with n the tanks of the dock are rapidly pumped out by powerful pumps located within the dock walls and the ship is lifted out of the water The water within the well of the floating dock spills out of the open ends of the dock The sectional floating docks may

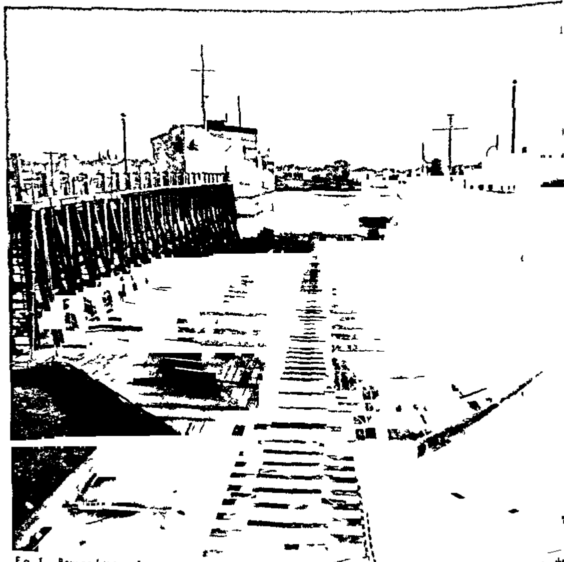


Fig 1 Bow-end view of a marine railway which has been pulled out of the water The tracks on which the rollers travel may be seen at right and left in the foreground There are also center tracks to take the load of the keel

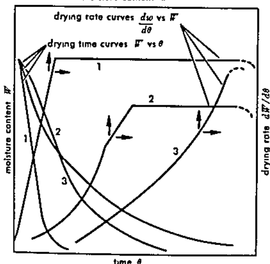


Fig 1 Drying time and drying rate curves illustrating the general problem of drying 1, Curves typical of a layer of thin material with most of the drying in the constant rate 2, A more general case where two stages in the falling rate period occur Typical of granular materials 3 A case in which no constant rate occurs Typical of homogeneous and colloidal materials such as soap, gelatin and viscous solutions

effects of the external drying medium such as air velocity, humidity, temperature and wet material shape and subdivision are studied with respect to their influence on the drying rate The results of such investigations are usually presented as drying rate curves and the natures of these curves are used to interpret the drying mechanism Figure 1 shows a series of typical drying rate curves

The constant rate period of drying when heat is supplied by convection is susceptible to theoretical and analytical treatment because it is essentially independent of the solid material When drying is accomplished by heat transfer from hot gases which also remove the evolved vapors the constant rate may be expressed in terms of heat transfer rates or mass transfer rates

A constant rate of evaporation at the surface of the solid maintains the surface at a constant temperature which in the absence of other heat effects is very nearly the wet bulb temperature of the air This temperature may range from 70 to 130°F for convection drying depending on the temperature and humidity of the air and on radiation This so-called wet bulb cooling effect is one reason why heat sensitive solids can be dried in air at temperatures well above the decomposition temperature of the solid

The magnitude of the constant rate can vary widely depending on the degree of subdivision of the material that is the manner in which the material is exposed to the drying air Thus the rate of drying in spray dryers can be several hundred thousand fold greater than the rates in tra

A number of empirical expressions experimental studies have been dev

timating the constant rate for different physical configurations of the wet material

When materials are dried in contact with hot surfaces termed indirect drying the air humidity and air velocity may no longer be significant factors controlling the rate Instead the "goodness" of contact between the wet material and the heated surfaces and the surface temperature will be controlling This may involve agitation of the wet material in some cases

The falling rate period is not as amenable to treatment as the constant rate period because the falling rate depends largely on the internal structure of the solid and the mechanism of moisture flow therein In falling rate processes the rate of drying decreases gradually until the moisture content of the material approximates the equilibrium value The equilibrium moisture content of a material is that moisture content to which a given material can be dried under specific conditions of air temperature and humidity A typical equilibrium moisture curve is shown in Fig 2 A unique characteristic of hygroscopic materials is that they hold or retain water at a vapor pressure less than water at the same temperature Moisture so retained is termed bound moisture Materials in which water exerts its normal vapor pressure at all moisture contents are termed nonhygroscopic, and in general are easier to dry

Classification of dryers Drying equipment for solids may be conveniently grouped into three classes on the basis of the method of transferring heat for evaporation The first class is termed di

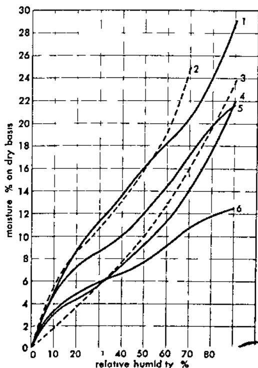


Fig 2 Equilibrium moisture content of organic materials at 70°F 1 lea soap; 4 wood 5 catgut 6,

- **All types of dryers used for producing a dry solid product from a wet feed**
  - **Direct Dryers**
    Heat transfer for drying is accomplished by direct contact between the wet solid and hot gases. The vaporized liquid is carried away by the drying medium that is the hot gases. Direct dryers might also be termed convection dryers.
    - **Continuous**
      Operation is continued without interruption as long as wet feed is supplied. It is apparent that any continuous dryer can be operated intermittently or batchwise if so desired.
      - **Direct Continuous Types**
        1. Continuous tray dryers such as continuous metal belts, vibrating trays utilizing hot gases, vertical turbo-dryers.
        2. Continuous sheeting dryers. A continuous sheet of material passes through the dryer either as festoons or as a taut sheet stretched on a pin frame.
        3. Pneumatic conveying dryers. In this type drying is often done in conjunction with grinding. Material conveyed in high temperature high velocity gases to a cyclone collector.
        4. Rotary dryers. Material is conveyed and showered inside a rotating cylinder through which hot gases flow. Certain rotary dryers may be a combination of indirect and direct types, for example hot gases first heat an inner shell and then pass between inner and outer shell in contact with the wet solid.
        5. Spray dryers. Dryer feed must be capable of atomization either by a centrifugal disk or a nozzle.
        6. Through circulation dryers. Material is held on a continuous conveying screen and hot air is blown through it.
        7. Tunnel dryers. Material on trucks is moved through a tunnel in contact with hot gases.
    - **Batch**
      Dryers are designed to operate on a definite size batch of wet feed for given time cycles. In batch dryers the conditions of moisture content and temperature continuously change at any point in the dryer.
      - **Direct Batch Types**
        1. Batch through-circulation dryers. Material held on screen bottom trays through which hot air is blown.
        2. Tray and compartment dryers. Material supported on trays which may or may not be on removable trucks. Air blown across material on trays.
  - **Infrared or Radiant Heat Dryers**
    Dielectric Heat Dryers
    The operation of radiant heat dryers depends on the generation, transmission and absorption of infrared rays.
    Dielectric heat dryers operate on the principle of heat generation within the solid by placing the latter in a high frequency electric field.
  - **Indirect Dryers**
    Heat for drying is transferred to the wet solid through a retaining wall. The vaporized liquid is removed independently of the heating medium. Rate of drying depends on the contacting of the wet material with hot surfaces. Indirect dryers might also be termed conduction or contact dryers.
    - **Continuous**
      Drying is accomplished by material passing through the dryer continuously and in contact with a hot surface.
      1. Cylinder dryers for continuous sheets such as paper, cellophane and textile piece goods. Cylinders are generally steam heated and rotate.
      2. Drum dryers. These may be heated by steam or hot water.
      3. Screw conveyor dryers. Although these dryers are continuous, operation under a vacuum is feasible. Solvent recovery with drying is possible.
      4. Steam tube rotary dryers. Steam or hot water can be used. Operation on slight negative pressure is feasible to permit solvent recovery with drying if desired.
      5. Vibrating tray dryers. Heating accomplished by steam or hot water.
      6. Special types such as a continuous fabric belt moving in close contact with a steam heated platten. Material to be dried lies on the belt and receives heat by contact.
    - **Batch**
      Batch indirect dryers are generally well adapted to operate under vacuum. They may be divided into agitated and nonagitated types.
      1. Agitated pan dryers. These may operate atmospherically or under vaccuum and can handle small production of nearly any form of wet solid that is liquids, slurries, pastes, or granular solids.
      2. Freeze dryers. Material is frozen prior to drying. Drying in frozen state is then done under very high vacuum.
      3. Vacuum rotary dryers. Material is agitated in a horizontal stationary shell. Vacuum may not always be necessary. Agitator may be steam heated in addition to the shell.
      4. Vacuum tray dryers. Heating done by contact with steam-heated or hot water heated shelves on which the material lies. No agitation involved.

Fig. 3. Classification of dryers based on methods of heat transfer

rect dryers, the second class indirect dryers and the third class radiant heat dryers In the chart in Fig 3 each class is subdivided into batch and continuous Batch dryers are restricted to low capacities and long drying times Most industrial drying operations are performed in continuous dryers The large numbers of different types of dryers reflect the efforts to handle the large numbers of wet materials in ways which result in the most efficient contacting with the drying medium Thus filter cakes pastes and similar materials when preformed in small pieces can be dried many times faster in continuous through circulation dryers than in batch tray dryers Similarly materials which are sprayed to form small drops as in spray drying dry much faster than in through circulation drying

Direct dryers The general operating characteristics of direct dryers are (1) drying is accomplished by convection heat transfer between the wet solid and a hot gas the latter removing the vaporized liquid as well as supplying the heat needed for evaporation (2) the heating medium may be steam heated air gases of combustion a heated inert atmosphere such as nitrogen or a superheated vapor such as steam (3) drying temperatures may range from prevailing atmospheric temperatures to 1400°F (4) at drying temperatures below the boiling point of the liquid increasing amounts of vapor of this liquid in the drying gas will decrease the rate of drying and increase the final liquid content of the solid (5) when the drying temperatures are above the boiling point throughout the process an increase in the vapor content of the gas or air in general will have no retarding effect on the drying rate and no effect on the final moisture content (6) for low temperature drying dehumidification of the drying gas is often required when high atmospheric humidities prevail (7) the efficiency of direct dryers will increase with an increase in the inlet temperature of the drying gas at a fixed exhaust temperature (8) the range of operating costs of direct continuous dryers will be in the range of $0 0005-0 025 per lb of dry product These figures include labor fuel power maintenance and depreciation For batch direct dryers the operating costs are generally considerably higher

Indirect dryers The general operating characteristics of indirect dryers are as follows (1) Drying by the transfer of heat by conduction and some radiation to the wet material conduction usually occurs through a metallic retaining wall The source of heat is generally condensing steam but may also be hot water gases of combustion molten heat transfer salts hot oil or electric heat (2) The drying temperature of the surface of contact may range from below freezing to 1000°F (3) Indirect dryers are especially suited to drying under reduced pressures and with inert atmospheres and are therefore well adapted to the recovery of solvents (4) Indirect dryers using condensing steam usually have a high efficiency because heat is supplied according to the demand but as in all dryers the efficiency falls off markedly when very low final moisture contents are required (5) Indirect dryers can handle dusty materials more readily than direct dryers (6) The operation of indirect dryers is frequently characterized by some method of agitation to improve the contact between the hot metal surface and wet material The nature of this contact determines the over all drying rate of indirect dryers heavy granular materials generally give higher heat transfer coefficients of contact than fluffy bulky solids

Radiant energy dryers These operate by the transfer of heat from a radiant source to the wet material being dried The temperature of the radiant source may range from hot water or steam temperatures 200-350°F to the temperatures of incandescent surfaces 1500-2500°F The generating medium may be steam hot liquids gas flames or electricity depending on the temperature desired and the equipment design

Special types Dielectric heat dryers do not fall in any of the above classes inasmuch as their operation depend on the generation of heat by high frequency fields inside the material being dried so that heat will actually flow out from the interior of the solid These dryers are used to dry large bulky objects which have a long internal path for moisture flow

Direct batch dryers In operation heated air circulates over the wet material being dried The wet solid is supported according to its physical form Lumber ceramics and similar massive objects are stacked in piles or on racks textile skeins painted objects and hides are suspended on hangers and granular materials pastes slurries and liquids are placed in trays which may be supported on stationary or movable racks Good performance of this type of dryer depends on uniform equal air velocities across all the wet material

In batch through circulation dryers heated air is blown through the wet material on screen bottom trays instead of across The material to be dried must be permeable to air flow

Dryers of this type but of binlike design are used extensively in the explosives industry to dry gunpowder and in food processing to dry and condition certain foodstuffs such as grains and corn

Direct continuous dryers Tunnel and continuous tray dryers usually consist of long enclosed housings or tunnels through which wet material is moved on trucks Hot air is blown through the trucks Air flow may be parallel counter flow or at right angles (cross flow) to the movement of the trucks The trucks may move continuously or semicontinuously through the tunnel Tunnel dryers may be operated adiabatically that is without the addition of heat in the tunnel or the air may be reheated periodically during its passage through the tunnel

Wet granular materials are held on the trays of trucks foodstuffs rayon cakes pottery and large ceramic objects are held on racks textile skeins are draped over rods and hides are pasted on s plates or hung on frames

In continuous through circulation dryers heated air is blown through a permeable bed of wet material as it passes continuously through the dryer Drying rates are much higher than in the usual tray or tunnel dryers because of the large surface area exposed per unit weight of material and because the smaller particles offer less resistance to internal moisture flow

The operation of this type of dryer depends on whether or not the wet material is in a state of subdivision suitable for through circulation of the hot air Some materials are already in such a permeable state Many materials require special preliminary treatment termed preforming to form them into permeable beds Preforming may include the processes of scoring on rotary filters granulating extruding briqueting flaking and predrying on finned drums

One type of through circulation dryer consists of a horizontal conveying screen which moves through a tunnel like housing A permeable bed of the wet material is supported and conveyed on the screen and hot air or gas is blown vertically up or down through the bed (Fig 4) Through circulation drying may also be performed in rotary type dryers which convey the material by a tumbling action imparted by the rotation of the dryer shell

Conveying screen through circulation dryers are widely used in chemical plants to dry fibrous flaky and granular materials such as cotton linters rayon staple cellulose acetate silica gel sawdust and minerals materials that are well suited to through circulation of air without prior treatment They are also used extensively to dry materials such as starch pigments calcium carbonate insecticides dyes and intermediates which must be preformed by one of the methods noted above

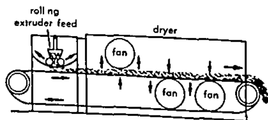


Fig 4 Path of travel of permeable bed through a three unit through-circulation dryer (*Proctor and Schwartz Inc*)

Direct rotary dryers are used extensively in the chemical industry A rotary dryer consists of a cylinder slightly inclined to the horizontal and rotated on suitable bearings The rotary action of the cylinder serves to convey wet material from one end to the other while passing hot gases axially through the dryer shell The contact of the solids and gases is further improved by means of flights arranged within the dryer shell so as to shower the wet material through the hot gas stream A rotary dryer without such lifting flights is usually called a rotary kiln

Rotary dryers may operate with air flow either parallel or countercurrent to the flow of the wet solid (Fig 5)

Pneumatic conveying dryers sometimes termed flash or dispersion dryers operate on the basis of simultaneously conveying and drying a wet solid in a high velocity stream of hot gas Temperatures up to 1400°F are used in these dryers The short contact times involved permit using gas temperatures above the decomposition temperature of the material The gas stream acts as both the conveying and heating medium Gas velocities on the order of 75 ft sec are used Frequently the wet material is in such a form that some disintegrating action is required before it can be conveyed A schematic diagram of this type of dryer is shown in Fig 6 It is applicable to granular free-flowing materials such as coal whey and sodium chloride and to sludges filter press cakes and similar nongranular solids which require disintegration for proper dispersion It is common practice to recycle dry product into the wet feed to facilitate dispersion and handling

Spray dryers operate on the principle of creating a highly dispersed liquid state in a high temperature (up to 1400°F) gas zone The heart of a spray drying process is the creation of small liquid droplets by spraying This may be accomplished by means of (1) high pressure nozzles (2) pneumatic nozzles and (3) high speed rotating disks Almost any pumpable liquid from a thin clear liquid to a pasty sludge can be atomized sufficiently for spray drying Liquids above a viscosity of 1500 centipoises however are very difficult to atomize and spray dry Generally fine atomization will not produce a large percentage of droplets less than 5 microns (μ) in diameter The particle size under

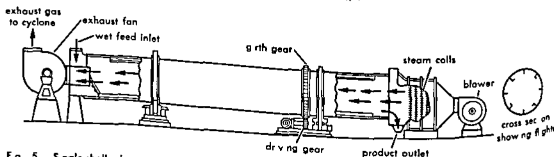


Fig 5 Single shell direct rotary dryer using steam heated air and balanced pressure by means of a blower and exhauster (*Hardinge Co*)

conditions of so-called coarse atomization will be on the order of 200–600 μ. Because of the high surface-volume ratio of small drops, the actual drying time in spray dryers may be considerably less than one second for high temperature operation.

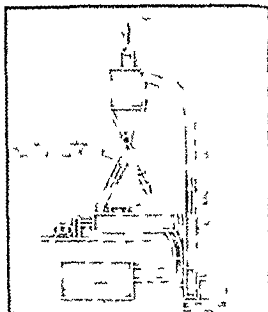


Fig. 6. Dispersion dryer.

Spray dryers are made in a multitude of designs. The drying gases may flow cocurrent with or countercurrent to the spray, and the spray may be directed vertically up or down or horizontally. Various types of spray dryers are shown in Fig. 7. They are used extensively in the chemical, pharmaceutical, and food industries.

Fluid bed dryers operate by having heated air pass upward at sufficient velocity through a column or layer of granular wet material to cause it to fluidize and become mixed by turbulent action. Wet feed may be introduced at the bottom of the column and dry product removed from the top, as shown in Fig. 8.

In direct continuous-sheeting dryers, heated air is circulated over or through a continuously moving wet sheet, which is supported by a variety of methods.

Direct continuous sheeting dryers occur in a variety of types. The festoon or loop dryer permits drying of continuous sheet material in a relaxed state by festooning or looping it over rolls, which in turn are conveyed through the dryer. Tenter dryers are used to dry a continuous sheet of material under tension. Heated air is usually blown perpendicularly from slots or nozzles against both sides of the sheet.

Direct continuous sheeting dryers are used to dry sheet materials that are sufficiently strong me

Fig. 7. (a) Cocurrent spray dryer. (b) Counterc

Fig 8 Fluid bed dryer

chanically in the wet and dry states to support their own weight or withstand tension as *required* Their widest use is in the textile field in the manufacture of coated and impregnated fabrics and in the preparation of films and coated papers

Indirect batch dryers *Vacuum shelf* dryers are indirect batch dryers which generally consist of a vacuum tight cubical or cylindrical chamber of cast iron or steel plate heated supporting shelves inside the chamber a vacuum source and a condenser In operation heat transfer takes place by conduction through metal surfaces and interfaces to the wet material held on the shelves

Vacuum shelf dryers are used extensively for drying pharmaceuticals temperature sensitive or easily oxidizable materials and small batches of high cost products where any product loss must be avoided This type is particularly useful for recovery of valuable solvents or vapors

Vacuum freeze dryers are used principally for drying materials that would be destroyed by the loss of volatile ingredients or by drying temperatures above the freezing point The material is dried in the frozen state so that a process of sublimation is involved that is ice sublimes directly to water vapor Because the material dries in a rigid frozen condition shrinkage is minimized and the resulting structure of the dry solid is usually porous and readily soluble This led to the term *lyophilization* in the early development of freeze drying

The equipment required for this method of drying consists of a heated drying chamber where sublimation occurs a piping system for the transport of vapors and a vapor removal system for condensable and noncondensable gases Removal of the condensable vapors may be accomplished by freezing on cold surfaces, by absorption in liquid desiccants or by adsorption on solid desiccants

One application of freeze drying (the Cryochem process) involves conduction heat transfer to the frozen solid held on a metallic surface However should the metal surface temperature rise above the freezing point of the solid melting will occur A second method of freeze drying utilizes heat transfer by radiation

Agitated pan dryers consist of a bowl or pan shaped receptacle in which wet material is stirred or agitated in contact with hot surfaces The operation may be atmospheric or vacuum Agitated pan dryers are used to handle small batches of materials that can withstand agitation during drying They are suitable for pastes and slurries and for materials containing valuable solvents which must be recovered

The vacuum rotary dryer is another type of batch indirect dryer with agitation It consists of a horizontal shell in which agitator blades attached to a horizontal rotating shaft revolve and agitate the material being dried (Fig 9) Heat is supplied by condensing steam in a jacket surrounding the shell, by hot water or by other heat transfer fluids Vacuum rotary dryers are used for large batches of materials that must be dried in the absence of air or where the recovery of solvents is required

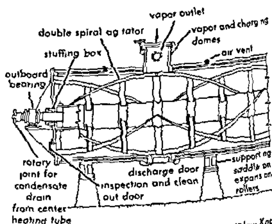


Fig 9 A typical vacuum rotary dryer (Blaw-Knox Co)

Indirect continuous dryers Indirect rotary dryers are similar mechanically and in appearance to the direct rotary dryers discussed above They differ primarily in that heat is transferred to the material through the metal shell or from steam tubes located around the dryer shell rather than from hot gases as in the case of direct rotary dryers

A screw conveyor dryer is essentially a jacketed conveyor in which material is heated and dried as it is conveyed In one type the jacket may extend only to the top of the conveyor which is left open to the atmosphere This is termed a trough dryer When the jacket encloses the conveyor completely a slight negative pressure is required to sweep out the evaporated moisture

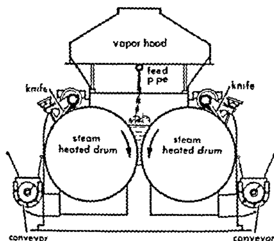


Fig 10 Double-drum dryer with pipe feed (*Buflovak Equipment Div , Blaw Knox Co*)

Vibrating conveyor dryers consist of a vibrating heated solid deck over which the wet solid moves This deck can be heated by hot gases steam methanol vapors or Dowtherm vapors which pass through a jacket fastened to the deck upon which the material is conveyed Direct gas flames can also be used as the source of heat A hood equipped with an exhaust fan and placed over the deck removes the evaporated liquid Infrared lamps may be mounted above the deck to increase drying rates

Drum drying consists of applying a liquid material solution slurry, or paste to a revolving heated metal drum which conducts heat to the wet film to evaporate the water during a partial revolution of the drum (Fig 10) The dry material is scraped from the drum by a stationary knife Drum dryers may be designated by type as atmospheric double drum atmospheric single drum atmospheric twin drum vacuum single-drum and vacuum double drum dryer

In drum drying the product is exposed to heat for only short periods of time This has the advantage that although the product may approach the temperature of the drum surface there usually is no adverse effect from overheating

Cylinder dryers sometimes called can dryers or drying rolls differ from drum dryers in that they are used for materials in a continuous-sheet form Cylinder dryers may consist of one large cylindrical drum such as the so-called Yankee dryer more often they comprise a number of drums arranged so that a continuous sheet of material may pass over them in series Typical of this arrangement are Fourdrinier paper machine dryers cellophane dryers and slashers for textile piece goods and fibers

The size of commercial cylinder dryers covers a wide range The individual rolls may be 2 6 ft in diameter and up to 20 ft in width In some cases the width of the rolls decreases throughout the dryer in order to conform to the shrinkage of the sheet

Drying of gases The removal of 95 100% of the water vapor in air or other gases is frequently necessary Gases having a dew point of −40°F are considered commercially dry The more important reasons for the removal of water vapor from air are (1) comfort as in air conditioning (2) control of the humidity of manufacturing atmospheres (3) protection of electrical equipment against corrosion short circuits and electrostatic discharges (4) requirement of dry air for use in chemical processes where moisture present in air adversely affects the economy of the process (5) prevention of water adsorption in pneumatic conveying and (6) as a prerequisite to liquefaction

Gases may be dried by the following processes (1) absorption by use of spray chambers with such organic liquids as glycerin or aqueous solutions of salts such as lithium chloride and by use of packed columns with countercurrent flow of sulfuric acid phosphoric acid or organic liquids (2) adsorption by use of solid adsorbents such as activated alumina silica gel or molecular sieves (3) compression to a partial pressure of water vapor greater than the saturation pressure to effect condensation of liquid water (4) cooling below dew point of the gas with surface condensers or cold water sprays and (5) compression and cooling in which liquid desiccants are used in continuous processes in spray chambers and packed towers—solid desiccants are generally used in intermittent operation that requires periodic interruption for regeneration of the spent desiccant

Desiccants are classified as solid adsorbents which remove water vapor by the phenomena of *surface adsorption* and *capillary condensation* (silica gel and activated alumina) solid absorbents which remove water vapor by chemical reaction (fused anhydrous calcium sulfate lime and magnesium perchlorate) deliquescent absorbents which remove water vapor by chemical reaction and dissolution (calcium chloride and potassium hydroxide) or liquid absorbents which remove water vapor by absorption (sulfuric acid lithium chloride solutions and ethylene glycol)

The *mechanical* methods of drying gases compression and cooling and refrigeration are used in large scale operations and generally are more expensive methods than those using desiccants Such mechanical methods are used when compression of the gas is a necessary step in the operation or when cooling of the gas is required

Liquid desiccants (concentrated acids and organic liquids) are generally liquid at all stages of a drying process Soluble desiccants (calcium chloride and sodium hydroxide) include those solids which are deliquescent in the presence of high concentrations of water vapor

Deliquescent salts and hydrates are generally used as concentrated solutions because of the practical difficulties in handling replacing and regenerating the wet corrosive solids The degree of drying possible with solutions is much less than the corresponding solids but, where only

ately low humidities are required and where large volumes of air are dried solutions are satisfactory See DESICCANT EVAPORATION FILTRATION GAS ABSORPTION OPERATIONS HEAT TRANSFER HUMIDIFICATION UNIT OPERATIONS VAPOR PRESSURE [W R M]

Bibliography E A Flosdorf and F J Stakes Freeze drying as applied to penicillin blood plasma and orange juice *Chem Eng Progr* 43 343–318 1947 W R Marshall Jr *Atomization and Spray Drying* 1954 J H Perry (ed) *Chemical Engineers Handbook* 3d ed 1950

Drying oil

Oils are classified as nondrying semidrying or drying according to their ease of autoxidation and polymerization to form a hard dry film on exposure to air More than 800 000 000 lb of drying oils is used annually in the United States in paints and varnishes Drying oils are relatively highly unsaturated that is they are composed of triglycerides constructed from unsaturated fatty acids The best drying oils contain several nonconjugated double bonds per molecule thus a good drying oil should have a high iodine value (above 130) should on hydrolysis yield only a small percentage of saturated acids (palmitic and stearic) and should furnish on hydrolysis large percentages (above 65%) of combined unsaturated acids such as oleic linoleic linolenic licanic and eleosteric acids See FAT AND OIL NONEDIBLE VINYLOGY

The table gives the percentage content as saturated or unsaturated glycerides of the common drying oils

Glycerides present in drying oils, %*

Name	Saturated	Oleate	Linoleate	Linolenate	Eleostearate	Licanate
Cottonseed	25	40	35			
Soybean	14	26	52	8		
Dehydrated castor oil	5	10	85			
Linseed	10	18	17	55		
Perilla	7	14	16	63		
Tung	5	7	3		85	
Oiticica	10	6	10			74

* Based on N A Lange (ed) *Handbook of Chemistry* 9th ed McGraw Hill 1956

Raw drying oils that is untreated drying oils are not suitable for paints and varnishes because they polymerize too slowly and various methods have been introduced to improve the polymerization process One method involves boiling the oil after addition of soluble *resin acid* salts of cobalt manganese or lead (such salts are known as driers) the pr d [illegible]
prox [illegible]
dries [illegible]
raw [illegible]
varnishes and enamels *Blown oil* is produced by blowing air through the oil (to which driers have been added) at about 120°C it is said to have superior wetting or surface covering properties Stand oil has been partially polymerized without admixture of driers by heating to 260–280°C This material is used extensively in antifouling paints, printing inks and linoleum as well as in varnishes and enamels Linseed oil is the most widely used drying oil in paints and varnishes See DRIER (PAINT), LINOLEUM, PAINT [E B R]

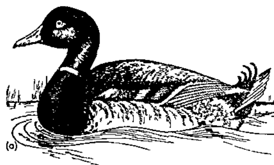
Duck

Any of more than 100 species of waterfowl belonging to the family Anatidae The family contains 166 species including the geese and swans Of the 46 Anatidae in the United States 35 are called ducks The ducks are commonly placed in five separate subfamilies They all have straight lamellated bills a fleshy tongue with fringed edges and webbed feet All are semiaquatic, they nest either on the ground or in trees but spend most of their time on or near water Many of them feed on land part of the time They are all migratory See GOOSE SWAN

The tree ducks subfamily Dendrocygninae a small group of only eight species are found in the warmer parts of the world There are two species in the United States both southern *Dendrocygna autumnalis* the black bellied tree duck nests from Texas south into Brazil The fulvous tree duck *D bicolor*, nests along the southern border of the United States from Louisiana to Texas and south into Argentina This duck is also found in India East Africa and Madagascar a most unusual distribution In Louisiana and Texas the bird is commonly called the squealer

Most abundant in number both of individuals and of species are the surface-feeding ducks or river and pond ducks of the subfamily Anatinae These birds do not dive but feed by dipping and consequently frequent shallow water They spring directly from the water when taking wing Virtually all of them are further characterized by the presence of a speculum a rectangular patch of metallic color near the posterior edge of the wing Interspecific hybrids among birds are rare but several instances are known among the members of this subfamily Most museums have examples of this crossing in their collections

Most common of all ducks is the mallard *Anas platyrhynchos* the greenhead familiar to hunters It is a Holarctic species breeding over most of Canada and in the United States in favored localities Flocks of semidomesticated mallards are commonly kept Others of this family that are well known include the teals black duck baldpate shoveller wood duck and pintail The latter species *Anas acuta* is another common duck It has been able to maintain its numbers because it is unusually wary and because its major breeding grounds are in the more remote parts of Alaska and northwestern Canada It is also a Holarctic species The wood duck *Aix sponsa* however has been seriously depleted in recent years and is only occasionally on the list of ducks which may be legally hunted This bird responds well to artificial nesting boxes and has been restored in numbers to









Wild ducks (a) The mallard duck, *Anas platyrhynchos*, length to 28 in (b) The American pintail *Anas acuta* length (drake) to 30 in (c) Canvasback duck *Aythya valisineria* length to 2 ft (d) The red breasted merganser, *Mergus serrator*, length to 25 in (*From E L Palmer Fieldbook of Natural History, McGraw Hill, 1949*)

some extent by protection and by the increased use of these boxes It will breed anywhere in southern Canada and the United States where it finds favorable conditions Many people think this is the most beautiful of all waterfowl

The subfamily Aythyinae includes the diving ducks or bay and sea ducks The diving duck is the better designation, because many species such as the redhead and canvasback, are found on interior waterways All dive for their food and commonly form large flocks, called rafts, in open deep water When taking wing they must run across the water to get underway The redhead, *Aythya americana*, whose breeding area is concentrated in the interior prairie lakes of southern Canada and the north central United States is an interesting species Its migration is not so much north and south as it is east and west and the redhead's favorite wintering ground is the Chesapeake Bay area Redheads lay large numbers of eggs in dump nests which are untended They also commonly lay their eggs in the nests of other ducks The canvasback, *Aythya valisineria*, is a prized food duck closely related to the redhead and frequents the same nesting, migratory, and wintering areas Both species are somewhat depleted because of both habitat destruction and overshooting

Other important diving ducks include the scaups, (called bluebills by most hunters) the scoters and eiders The beautiful harlequin duck, *Histrionicus histrionicus*, of the western United States is also a diving duck

The subfamily Erismaturinae, the ruddy and masked ducks are represented in the United States by only one species, the small active ruddy duck, *Oxyura jamaicensis* Its short erect tail and chestnut breast identify it immediately It is so easily shot that it is often called fool duck

Another of the subfamilies is the Merginae, the mergansers These are the fish eating ducks whose bills are slender and are equipped with toothed edges as a fish catching modification They are not favored by hunters because of their fishy flavor but are frequently shot as a target bird and sometimes by fishermen who believe that the mergansers are detrimental to fishing especially in trout streams There is little indication that they have any significant relation to fish populations These are beautiful birds, of which three species occur in the United States

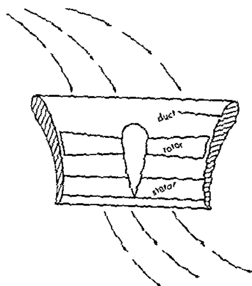
For a discussion of domesticated barnyard ducks *see* POULTRY PRODUCTION [J D B]

Ducted fan

A propeller or multibladed fan inside a coaxial duct or cowling also called a ducted propeller or a shrouded propeller, although in a shrouded propeller the ring is usually attached to the propeller tips and rotates The duct serves to protect the fan blades from adjacent objects and to protect them from the revolving blades but more importantly, the duct prevents radial flow of the fluid at the blade tips Fan efficiency remains high over a wider speed range with a properly shaped duct than without However fan efficiency is sensitive to duct shape at off center design conditions Without a well rounded inlet lip and a variable area exit, off center performance may be worse than without the duct

With a duct static thrust for a given power input is higher than without one For this reason propellers of vertical take off and landing aircraft may be ducted At low speeds, a stator of radial airfoils downstream from the propeller or an

oppositely turning coaxial propeller improves efficiency by converting slipstream rotation into axial velocity The duct may also form a nozzle to further increase exit jet velocity Air flow past the outer contour of the duct influences over all performance

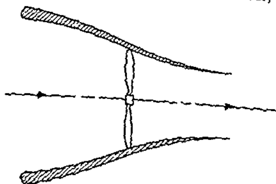


Static thrust and operating speed range of fan is improved by a duct

Ducted fans are used in axial flow blowers or compressors of several stages for turbine engines (see GAS TURBINE TURBOFAN) In such [illegible] of air than passes through the turbine the additional air leaving at lower exit velocity and hence higher jet propulsion efficiency for moderate speed aircraft than obtainable with a simple turbojet See AIRCRAFT PROPULSION CONVERTIPLANE TURBOJET [F H R]

Ducted flow

Fluid flow with zero velocity at the boundary relative to the boundary This condition is distinguished from jet flow in which the boundary is a fluid either liquid or gas and does not remain stationary



Ducted flow past a propeller

Under certain conditions a propeller has a housing or shrouding around it as in the sketch, to control the fluid approaching it as well as to control the induction of fluid into the jet downstream from the propeller Shrouding a propeller may be used on a ship to decrease interference of propeller and hull or it may be used to protect the propeller Additional frictional losses are incurred by the high velocity flow near the surface of the ducts which in some instances may be offset by improved guidance and control of the jet [V L S]

Dumortierite

A nesosilicate mineral with composition $AlBSi_3O_?(OH)$ Dumortierite crystallizes in the orthorhombic system but well formed crystals are rare the mineral usually occurs in parallel or radiating fibrous aggregates There is one direction of poor cleavage The hardness is 7 on Mohs scale and the specific gravity is 3 26–3 36 The mineral has a vitreous luster and a color that varies not only from one locality to another but in a single specimen It may be pink, green blue or violet Dumortierite is found in schists and gneisses and more rarely in pegmatites In the United States it occurs at Dehesa Calif and at Rochester Nev, where it has been mined for the manufacture of high grade porcelain See SILICATE MINERALS [C S H[.]]

Dune

A mound or hill of wind blown debris usually sand Dunes commonly occur in desert or arid regions on or near sandy shores of lakes and oceans on or near river flood plains especially if the volume of water varies greatly, on or near the bare outwash plains of retreating glaciers, and in areas of severe drought where the covering vegetation has been removed Particles are borne or bounced along by the wind until decreasing velocity turbulence or an obstacle causes deposition The resultant deposit further breaks the force of the wind causes more deposition and thus accelerates its own growth If sand is abundant and the wind is strong dunes may attain heights greater than 100 ft or more rarely as in the Sahara of northern Africa heights greater than 400 ft See SEDIMENTATION (GEOLOGY), see also LOESS

Form and structure Dunes may occur as individual barchans (crescentic dunes), as transverse ridges or as longitudinal ridges Barchans are usually the result of unidirectional wind of moderate velocity and a small sand supply Horns of the crescent point away from the wind (Fig 1) Similar wind conditions and a larger supply of sand form transverse ridges which are at right angles to wind direction Strong winds of variable direction and large sand supply result in longitudinal ridges, elongated parallel to principal wind direction Seif (sword) dunes are longitudinal ridges that taper to a point

Complex dunes may completely blanket an area where there is strong wind turbulence and an abun

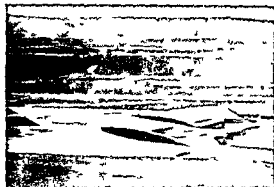


Fig 1 Barchans or crescent shaped dunes near Biggs Oregon These river bed dunes are moving from left to right with steep side away from the wind (USGS)

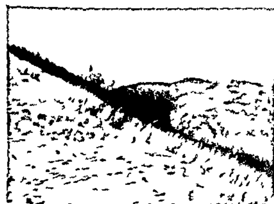


Fig 2 Steep advancing front of dune arrested by new plant growth Indiana Dunes State Park

dance of sand Dunes may be crescent shaped with the horns of *one crescent encroaching upon* the convex side of the next This results in a series of sand peaks and depressions or fuljis They may also occur as a complicated maze of intersecting sinuous ridges with intervening fuljis Broad expanses of sandy desert consisting either of continuous dunes or a sand sheet are called ergs *See* DESERT EROSION FEATURES

Most dunes consist of quartz sand mixed with feldspar mica clay minerals and fine rock fragments although dunes of clay silt gypsum and travertine are known Sizes of material may range from 1 μ to 2 mm [T C]

Migration of dunes Unless secured by vegetation dunes usually migrate in the direction of the prevailing wind Sand is blown up the windward slope and deposited on the steeper leeward slope as the dune advances (Fig 2) *Rate of migration* varies from a few feet to more than 100 ft per year In some areas forest trees and even buildings have been buried and then uncovered after a period of years Examples are found in the Indiana Dunes State Park at the south end of Lake Michigan on the Baltic coast of Prussia and along the shore of the Bay of Biscay northward from Bayonne

Along the eastern coast of United States—where shore processes have deposited and redeposited sands in the shape of sand bars spits hooks and barrier beaches the wind has taken the sands and piled them into dunes As the dunes continue to grow the youthful dunes constantly migrate while the older ones become anchored by vegetation Dunes of this type may be found on Cape Cod and the Outer Banks of North Carolina [N A B]

Dune vegetation Along the shores of the Great Lakes particularly Presque Isle in Lake Erie and the shores of Lake Michigan the stabilizing effect of vegetation on the shifting sand has been studied extensively As the root tangles bind the transported particles and eventually as sand accumulates a successional series of dunes and vegetation can be observed The plants which germinate on *the sandy beach are mainly xerophytes They must* withstand the effects of wind and drifting sand which either bury them or expose their rhizomes or roots Among the most successful plants associated with dune formation are those which propagate extensively by rhizomes such as sand reed (*Calamovilfa*) marram or beach grass (*Ammophila*) beard grass (*Andropogon*) and reed bent grass (*Calamagrostis*)

At Presque Isle the vegetational succession from beach to forest as well as the formation and stabilization of dunes can be studied in the space of a few hundred yards The sandy beach at the water s *edge is usually devoid of vegetation because of* wave action The first vegetation appears at the back beach about 75 ft from the water s edge This consists of seedlings of sea rocket (*Cakile*) beach pea (*Lathyrus*) and Jerusalem oak (*Chenopodium*) The first dune ridge is heavily populated with beach grasses predominantly *Ammophila* In the swale between the first and second ridges grasses commonly occur as well as a few seedlings of cottonwood (*Populus deltoides*) and willows (*Salix*) On the crest of the second dune about

means of a Swedish increment borer the age of these trees was determined as 10 or 11 years) On the lee side of the dune a few willows (*Salix cordata*) an occasional *Rubus* sorrel or dock (*Rumex*) and rock cress (*Arabis*) are found Continuing inward toward the areas of older more permanent vegetation there is a successional transition and stabilization of the dunes Poplars are the dominant species on the ridges and *Calamagrostis* horsetails (*Equisetum*) and sedges (*Carex Juncus*) are numerous in the swales between the ridges Beyond the dunes thickets of *Viburnum* shadbush *Juniper* and bayberry gradually *merge with pine and some oaks and eventually* with red oak and red maple this is the most mature forest on Presque Isle [C B C]

Bibliography R A Bagnold *The Physics of Blown Sand and Desert Dunes* reprint 1954 H C Cowles The ecological relations of the vegetation

on sand dunes of Lake Michigan *Botan Gaz* 27 95 116 167 202 281 308 361 391 1899 J T Hack Dunes of the western Navajo country Geog Rev 31 240–263 1941 F A Melton A tentative classification of sand dunes—its application to dune hi tory in the southern High Plains *J Geology* 48 113–145 1940 H T U Smith *Geological Studies in Southwestern Kansas* Kansas Geol Survey Bull 34 1940

Dust and mist collection

The physical separation and removal of particles either solid or liquid from a gas in which they are suspended Such separation is required for one or more of the following purposes (1) to collect a product which has been processed or handled in gas suspension as in spray drying or pneumatic conveying (2) to recover a valuable product inadvertently mixed with processing gases as in kiln or smelter exhausts (3) to eliminate a nuisance as in fly ash removal (4) to reduce equipment maintenance as in engine intake air filters (5) to eliminate a health fire explosion or safety hazard as in bagging operations or nuclear separations plant ventilation air and (6) to improve product quality as in cleaning of air used in processing pharmaceutical or photographic products Achievement of these objectives involves primarily gas handling equipment but the design must be concerned with the properties and relative amounts of the suspended particles as well as with those of the gas being handled

All particle collection systems depend upon subjecting the suspended particles to some force which will drive them mechanically to a collecting surface The known mechanisms by which such deposition can occur may be classed as gravitational inertial physical or barrier electrostatic molecular or diffusional and thermal or radiant There are also mechanisms which can be used to modify the properties of the particles or the gas to increase the effectiveness of the deposition mechanisms For example the effective size of particles may be increased by condensing water vapor upon them or by flocculating particles through the action of a sonic vibration Usually larger particles simplify the control problem To function successfully any collection device must have an adequate means for continuously or periodically removing collected material from the equipment

Devices for control of particulate material may be considered by structural or application similarities in eight categories

Gravity settling chamber In this the simplest type of device but not necessarily the least expensive the velocity of the gas is reduced to permit particles to settle out under the action of gravity Normally settling chambers are useful for remov

that the particles have greater inertia than the gas

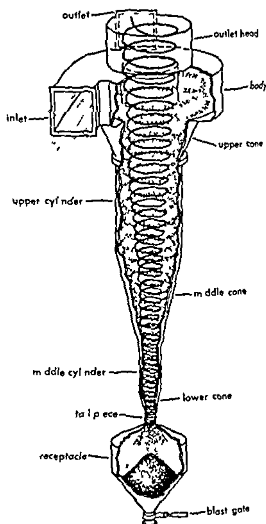


Fig 1 Cyclone dust separator (Amer can Blower Corp)

The cyclone separator, typical of this type of equipment is one of the most widely used and least expensive types of dust collector In a cyclone the gas usually enters a conical or cylindrical chamber tangentially and leaves axially Because of the change of direction the particles are flung to the outer wall from which they slide into a receiving bin or into a conveyor while the gases whirl around to the central exit port (Fig 1) A large variety of configurations are available For large air handling capacities an arrangement of multiple small diameter units in parallel is often used to attain high collection efficiencies and to permit lower headroom requirements than would be possible with a single unit

Mechanical inertial units are similar to cyclones except that the rotational motion of the gas is induced by the action of a rotating member Some such units are designed to act as fans in addition to their dust collecting function There are also a wide variety of other units many are called impingement separators Most separators used to remove entrained liquids from steam or compressed air fall into this category

Packed bed A particle laden gas stream may be cleaned by passing it through a bed or layer of packing composed of granular materials such as sand coke gravel and ceramic rings or fibrous materials such as glass wool steel wool and textile staples Depending on the application the bed depth may range from a fraction of an inch to several feet Coarse packings which are used at relatively high throughput rates (1 15 ft/sec superficial velocity) to remove large particles rely primarily on the inertial mechanism for their separating action Fine packings operated at lower throughput rates (1 50 ft/min superficial velocity) to remove relatively small suspended particles usually depend on a variety of deposition mechanisms for their separating effect Packed beds because of a gradual plugging caused by particle accumulation are usually limited in use to collecting particles present in the gas at low concentration unless some provision is made for removing the dust—for example by periodic or continuous withdrawal of part of the packing for cleaning Depending on the application and design the collection efficiencies of packed beds range widely (50–99 999%)

Cloth collector In such a collector also known as a bag filter the dust laden gas is passed through a woven or felted fabric upon which the gradual deposition of dust forms a precoat which then serves as a filter for the subsequent dust These units are analogous to those used in liquid filtration and represent a special type of packed bed Because the dust accumulates continuously the resistance to gas flow gradually increases The cloth must therefore be vibrated or flexed periodically

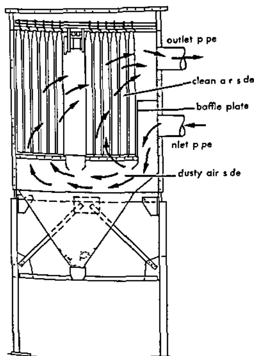


Fig 2 Cloth collector (*Wheelabrator Corp*)

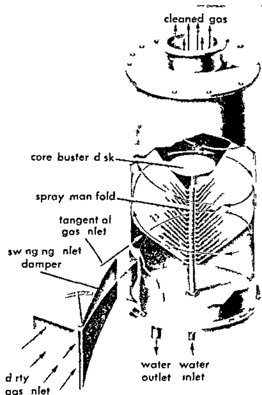


Fig 3 Cyclonic liquid scrubber (*Chemical Construction Corp*)

to dislodge accumulated dust (Fig 2) A wide variety of filter media is available Cotton or wool sateen or felts are usually used for temperatures below 212°F Some of the snythetic fibers may be used at temperatures up to 300°F Glass and asbestos or combinations thereof have been used for temperatures up to 650°F For special high tem

ter to assist in the particulate collection process An extremely wide variety of equipment is available ranging from simple modifications of corresponding dry units to permit liquid addition to devices specifically designed for wet operation only (Fig 3) When properly designed for a given application scrubbers can give very high collection efficiency although the mere addition of water to a gas stream is not necessarily effective

Electrostatic precipitator Particles may be charged electrically by a corona discharge and caused to migrate to a collecting surface The single stage unit which is commonly known as a Cottrell precipitator and in which the charging and collecting proceeds simultaneously is the type generally used for industrial or process applications These units normally employ direct current at voltages ranging from 30 000 to 1t vol

The two-stage unit in which charging and collection are carried out successively is commonly used for air conditioning applications. These units also employ direct current ranging from 5 000 to 13 000 volts and involve close internal clearances (0.25–0.5 in.). Electrical precipitators are capable of high collection efficiency of fine particles. The reentrainment of collected material as flocs in the exhaust gas, a phenomenon known as snowing, must be avoided to prevent a possible accentuated nuisance or vegetation damage problem. *See* ELECTROSTATIC PRECIPITATOR.

Air filter This is a unit used to eliminate very small quantities of dust from large quantities of air, as in air conditioning applications. Although units in this class actually fall into one of the previous classes, they are given a special category because of wide usage and common special features. In this category are viscous coated fiber mat filters and dry filters. These are actually a form of packed bed and are frequently known as unit filters; they are available as standard packaged units from a large number of manufacturers. The domestic furnace filter is an example of a viscous coated unit filter. Automatic filters provided with continuous and automatic cleaning arrangements are available in both the viscous coated and dry forms, as well as with electrostatic provisions.

Miscellaneous equipment Acoustic or sonic vibrations imparted to a gas stream cause particulates to collide and flocculate, forming larger particles that are more readily collected in conventional apparatus. This principle has been employed but has had extremely limited application because of economic and other practical considerations. In thermal precipitation, suspended particles are caused to migrate toward a cold surface or away from a heated surface by the action of a temperature gradient in the gas stream. This principle has found extensive use in atmospheric sampling work. *See* AEROSOL; AIR POLLUTION CONTROL; ATMOSPHERIC POLLUTION; PARTICLE PROPERTIES; SEPARATION (CHEMICAL AND PHYSICAL); SEPARATION (MECHANICAL); SMOKE; UNIT OPERATIONS. [C.E.L.]

Bibliography: R. E. Kirk and D. F. Othmer (eds.), *Encyclopedia of Chemical Technology*, vol. 7, 1951; C. E. Lapple, Elements of dust and mist collection, *Chem. Eng. Progr.*, 50(6):283–287, 1954; K. E. Lunde and C. E. Lapple, Dust and mist collection, *Chem. Eng. Progr.*, 53(8):385–391, 1957; J. H. Perry, *Chemical Engineers' Handbook*, 3d ed., 1950; U.S. Atomic Energy Commission, *Handbook on Air Cleaning*, 1952.

Dust storm

A strong, turbulent wind carrying large clouds of dust. In a large storm, clouds of fine dust may be raised to heights well over 10 000 ft and carried for hundreds or thousands of miles.

Sand storms differ by the larger mass, more rapid settling speeds of the particles involved, and by the stronger transporting winds required. The sand cloud seldom rises above 50–100 ft and is not carried far from the place where it was raised.

Dust storms cause enormous erosion of the soil, as in the dust bowl disasters of 1933–1937 in the Great Plains of the United States. Besides causing acute physical discomfort, they present a severe hazard to transportation by reducing the visibility to very low ranges. Conditions required are an ample supply of fine dust or loose soil, surface wind strong enough to stir up the dust, and sufficient atmospheric instability for marked vertical turbulence to occur.

Mechanics of dust raising Dust (or sand) is initially raised when particles become dislodged by aerodynamic stresses of the strong winds upon exposed grains. The larger particles fall obliquely after attaining considerable horizontal speed, bombarding other particles on the surface which in turn become dislodged and further the process.

R. A. Bagnold classifies as dust particles having diameters of 10^{-3}–10^{-2} mm with free fall speeds ranging 10 –5 cm/sec; sand particles 0.1–1 mm in diameter have fall speeds 10–100 cm/sec. Dust can be readily carried upward by ordinary turbulent eddies in an unstable air mass (*see* ATMOSPHERIC ADIABATIC CHANGE), but these are generally too feeble to sustain large sand particles, which attain only small heights in passing. In any dust storm, particles of various sizes are raised, the smallest being carried to greatest heights, from which they may take days or weeks to settle while remaining as a dust haze.

Soil factors. Soil condition is the most decisive criterion for development of dust storms. This depends on vegetative cover and upon binding of soil by moisture, both factors being dependent upon prior rain or snowfall. Dust storms are most frequent in spring, when in semiarid regions the earth is least covered by vegetation. Loosening of soil and overturning of humus by spring plowing and overgrazing of grasslands are prime contributors to setting up soil conditions favorable for dust storms.

Meteorological factors. Surface wind speed required vary according to soil characteristics. In some desert regions, sand storms occur with winds of 15–20 mph. Extensive dust storms in North America usually require winds of 25–30 mph or more. Such winds are present over large areas in the circulations of many well-developed cyclones, which may raise dense dust clouds several hundred miles across if soil conditions are right.

A further requisite is thermodynamic instability (strong decrease of temperature with height), necessary for development of the vertical eddies required to transport dust aloft from surface layers. Most dust storms occur in daytime, particularly in the afternoon, when the air is warmest at the ground, hence most unstable just above. Major dust storms in the United States are almost exclusively confined to maritime polar air masses from the Pacific Ocean, which are characteristically unstable to great heights (*see* AIR MASS).

Dust storms associated with large scale wind systems are also common in the Sahara and Gobi deserts More local but often severe dust storms resulting from thundersqualls are common in all the desert regions (*see* SQUALL)

Optical and electrical effects Small dust particles increase scattering of light mainly in short (blue) wavelengths The sun often appears a deep orange or red when seen through a dust cloud however optical effects are variable Large particles are effective reflectors and an observer in an aircraft above a dust storm may see a solid sheet with an apparent dust horizon

Due to friction with air or ground dust particles acquire appreciable electrostatic charges and on striking radio antennas may cause severe static Visible electrical discharges sometimes occur within the dust cloud [C W N]

Bibliography R A Bagnold *The Physics of Blown Sand and Desert Dunes* 1942 H R Byers *Synoptic and Aeronautical Meteorology* 1937

Dwarf star

The most *common stars in the* local galaxy are dwarf or main sequence stars The Sun is a typical dwarf with surface temperature of 5750°K radius of 690 000 km mass of 2×10^{33} g luminosity of 4×10^{33} ergs/sec (that is 4×10^{23} kilowatts) These numbers are fairly typical as to radius and mass However in luminosity main sequence stars occur in a range from O and B stars up to 10^4 times brighter than the Sun with surface temperatures of 35 000–15 000°K and down to the M dwarfs 10^{-4} as bright as the Sun with a temperature of only 2500°K Dwarfs form a single parameter family in which the significant variable is mass members range in mass from 20 Suns down to less than 0 1 Sun The M or red dwarfs are the most common stars in space and because of their low luminosity have the longest life Almost all other types of stars have evolved from main sequence stars *See* STAR [J L GR]

Dyadic

A mathematical abstraction corresponding to an expression of the type $\beta\gamma + \delta\epsilon +$ in which the elements (dyad symbols) consist of two vector symbols in juxtaposition without the intervention of either the dot (•) or cross(×) Essentially a dyad is an ordered pair of vectors subject to certain rules of operation The first symbolic factor in a dyad (β in $\beta\gamma$ for example) is called the antecedent and the second the consequent

The concept dyadic (J W Gibbs) is inherent in the formal structure of certain vectorial expressions such as $(\alpha \cdot \mathbf{i})\mathbf{i} + (\alpha \cdot \mathbf{j})\mathbf{j} + (\alpha \cdot \mathbf{f})\mathbf{f}$ and $(\alpha \cdot \beta)\gamma + (\alpha \cdot \delta)\epsilon$ To reduce $(\alpha \cdot \beta)\gamma + (\alpha \cdot \delta)\epsilon$ to the form $\alpha \cdot (\beta\gamma + \delta\epsilon)$ it is sufficient to accept $\beta\gamma + \delta\epsilon$ as a mathematical entity and to agree that whenever $\alpha \cdot$ is applied from the left to $\beta\gamma + \delta\epsilon$ it is to be distributed to the antecedents while the symbol $+$ is to become the plus of vector addition Similarly $(\beta\gamma + \delta\epsilon) \cdot \alpha = \beta(\gamma \cdot \alpha) + \delta(\epsilon \cdot \alpha)$ by distribution to the consequents

Two dyadics Φ and Ψ are by definition equal if and only if $\Phi \cdot \alpha = \Psi \cdot \alpha$ identically in α In particular if $\beta \neq 0$ $\gamma \neq 0$ and $\beta\gamma = \gamma\beta$ then $\beta(\gamma \cdot \alpha) - \gamma(\beta \cdot \alpha)$ and β and γ are parallel

If the $\mathbf{r}_a$ are the base vectors $\partial\mathbf{r}/\partial x^a$ of a space defined by $\mathbf{r} = \mathbf{r}(x^1 \quad x^n)$ and $\mathbf{r}^b - g^{bc}\mathbf{r}_c$ and the quantities $T_b{}^a$ are the components of a tensor of the type (1 1 w) then $T_b{}^a\mathbf{r}_a\mathbf{r}^b$ is an invariant Consequently tensors can be related to dyadics *See* CALCULUS OF TENSORS CALCULUS OF VECTORS [H V C]

Dye

A colored substance which imparts more or less permanent color to other materials *See* DYEING

Not all colored substances are dyes however If red iron rust is ground with white sugar the resulting mixture has an over all reddish appearance The sugar has been colored by pigmentation and the iron rust has been used as a pigment (*see* PIGMENT) Examination of the mixture under a microscope shows distinct white and red particles and a separation can be made by dissolving the sugar out of the mixture with water If the red iron rust is added to white cloth in water some of the red particles may cling to the cloth but they can be removed by rubbing or by washing with soap Some colored substances (usually organic chemical compounds) may be added to cloth in water and after a period of soaking usually accompanied by heat and agitation the cloth will be colored no separate colored particles can be seen under ordinary microscopic examination and the color cannot be removed by washing even with soap The cloth has been dyed and the colored substance is a dye (also called a dyestuff)

Customarily colored water insoluble substances are called pigments Dyes are generally water soluble although some are soluble only during application after which they become insoluble

The mechanism by which soluble colored substances enter the internal structure of fibers and there become fixed has been variously explained in terms of the physical and chemical concepts of the times when the explanations were given It is said to be an adsorption phenomenon a salt formation a quasi chemical union caused by hydrogen bonding or an ether linkage and in some cases it is considered to be a true solution effect The end result however is that the dye has imparted a color (not necessarily that of the solid dye itself) to the fiber which is more or less resistant to washing or removal by similar mechanical operations The dye is said to be fixed on and to have affinity or substantivity for the material it has colored The material is designated as the substrate If the color is quite resistant to washing and light it is called a fast color if the color is easily removed or fades quickly it is a fugitive dye

Because not all water soluble colored substances are dyes various attempts have been made to re

late chemical constitution with color and substantivity One of the earlier and still very useful explanations was given by O N Witt who stated in 1876 that all colored organic compounds (called chromogens) contain certain unsaturated chromophoric groups which are responsible for the color and if these compounds also contain certain auxochromic groups they possess dyeing properties Examples of chromophores are the groups —NO, —N—N and =CO and of auxochromes —NH, OH The auxochromes also influence hue according to their nature number and position on the chromogen molecule An elaboration of this theory in 1888 by H E Armstrong regarded all chromophores as being quinoid (=R) in structure Many later studies have added explanations of the nature and variation of color in organic compounds but the Witt theory provides a frame of reference for most dyes which is simple practical and satisfactory for all but the specialist in dye chemistry *See* SPECTROPHOTOMETRIC ANALYSIS

Dyestuffs may be classified in various ways according to color (blue red and so on) origin (natural from vegetable and animal matter or synthetic) chemical structure kinds of material to which they are applied and method of application

Color or hue Commercial dyes are usually *named to indicate the hue imparted to the dyed article This color is not necessarily the same as that of the solid dyestuff but it is the same as the color of the dye solution from which the dyeing is made*

Origin stuffs (saff rived from ious plant the extract of the Mediterranean mollusk *Murex brandaris* gave Tyrian purple so *expensive that it* was the mark of a king and *indigo came from* plants of the genus *Indigofera* Most natural dyes are of the mordant type *that requires a fixing* agent With the advent of the *synthetic dyes* which are far more var ed n sh de

their greatest use now is in the dyeing of leather

Synthetic dyes were an *early result* of the *devel*opment of chemistry as a science and the first commercial production in 1857 *in* England of a synthetic organic chemical *was that of* the first commercially produced *synthetic* dye mauve discovered in 1856 by Sir *William Henry Perkin* The early history of organic chemistry *was primarily* a record of investigations of *material* and methods for making synthetic dyes Because one of the *early* and important materials used *in these* syntheses was aniline (derived from coal *tar*) *the* synthetic dyes have been known as coal tar *dyes and* aniline dyes Their manufacture was at first so vigorously pursued in Germany that *it became* almost exclusively a German industry Only *in* World War I (1914–1918) did the loss of dye imports from Germany lead to the development of their manufacture in other countries Since that time manufacture has grown in the United States so that 54 manufacturers produced 143 000 000 lb in 1957

The manufacture of dyes proceeds from simple raw materials mostly aromatic hydrocarbons such as benzene (benzol) toluene and naphthalene with introduction of other chemical groups such as nitro amino halogen and sulfonic acid These intermediate compounds are then further processed by many special operations in organic chemistry such as diazotization coupling condensation and fusion to give the final dyestuff Following the chemical manufacture the dye may be treated to give it special physical properties and is standardized for strength It may be sold as a dry powder as a paste or in solution *See* DIAZOTIZATION

Of the many thousands of dyes which have been synthesized in laboratories about 3500 have had actual *commercial use* These have been indexed in several compilations the latest and most complete being the *Colour Index* 2d ed 1956 published jointly by the Society of Dyers and Colourists (England) and the American Association of Textile *Chemists and Colorists* This encyclopedic work is the major reference for the dye chemist Each mber as ufac com mon names with general usage commercial dyes are sold under a large number of names *many of* which relate to the same material Each dyestuff manufacturer has his own nomenclature which usually consists of three or four parts First comes his own trademarked name for the class then the hue then words letters or numbers which *describe the* shade and other characteristics and finally designation of the strength or physical *form* (powder paste and double strength powder) Thus Erie Black RX Conc Paste is a trade-name for the color (CI30235) which is a direct cotton dye *black with* a reddish shade of extra quality (*improved over* earlier manufacture) and standardized in strength much stronger than the ordinary type

Chemical structure The most precise and scientific classification of dyes is based upon their chemical structure This is the classification *of interest* to the research chemist and the manufacturer of dyes

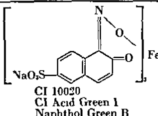
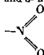
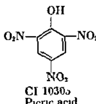
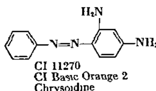
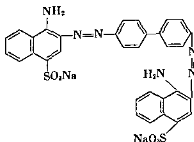
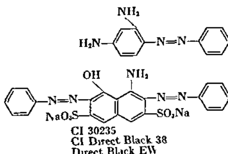
The table shows the 22 classes that are *listed in* the *Colour Index* with typical structures and examples

Utilization This classification is based *on the* materials to which the dyes are applied

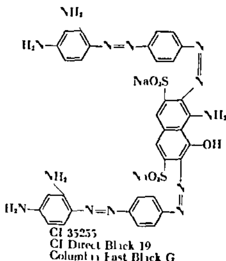
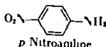
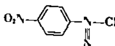
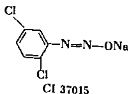
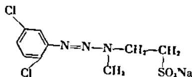
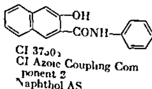
Cloth of natural fibers Coloring cotton wool linen and silk is by far the largest use for dyes

Cloth of synthetic fibers These include regenerated cellulose (viscose) cellulose acetate nylon polyacrylate and *polyester* Regenerated cellulose fiber can be dyed with dyes for cotton dyeing other synthetic fibers with the colors commonly used for

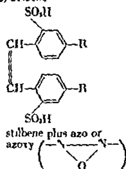
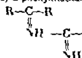
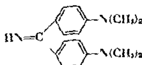
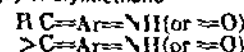
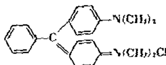
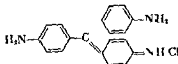
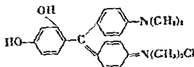
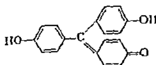
Some important dyes

Class basic chemical structure chromophore	Example	
(1) Nitroso (quinone oxime) o-Nitrosophenol (or o-nitrosonaphthol) —N=O	 CI 10020 CI Acid Green 1 Naphthol Green B	Mordant dyes used only as lakes of metals prepared by action of nitrous acid on phenols or naphthols
(2) Nitro o- and p-Nitrophenols and o- and p-nitroamines 	 CI 10305 Picric acid	Prepared by action of nitric acid on phenols naphthols and amines
(3) Azo Aromatic and heterocyclic azo compounds —N=N—		Most numerous class of dyes
(a) Monazo R—N=N—R	 CI 11270 CI Basic Orange 2 Chrysoidine	Prepared by action of nitrous acid on primary arylamine to give diazo compound which is coupled with aromatic amino or hydroxy compound
(b) Disazo R—N=N—R — N=N—R	 CI 22120 CI Direct Red 28 Congo Red	Prepared by nitrous acid on primary diamine with coupling of the resulting tetrazo compound with two mols of aromatic amino or hydroxy compound
(c) Trisazo R—N=N—R — N=N—R — N=N—R	 CI 30235 CI Direct Black 38 Direct Black EW	Prepared from four intermediates by rediazotizations and further couplings

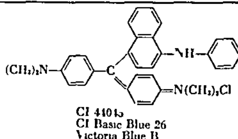
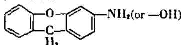
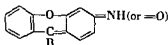
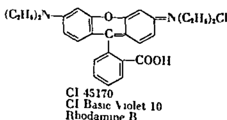
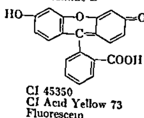
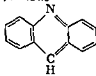
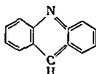
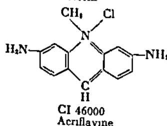
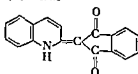
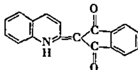
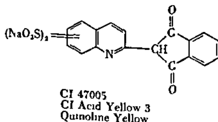
Some important dyes (Cont.)

Class, basic chemical structure, chromophore	Example	
(d) Polyazo With four or more azo groups	 CI 35255 CI Direct Black 19 Columbia Fast Black G	
(4) Azoic Azo compounds with —N=N— as chromophore formed on fiber from components		
(a) Fast-color base RNH_2	 *p*-Nitroaniline	Primary amine which may be diazotized
(b) Fast-color salt $RN{\equiv}N$ with Cl	 CI 37035 CI Azoic Diazo Compound 37 Diazotized *p*-nitroaniline Fast Red GG Salt	Diazo salt stabilized with additives or by formation of complex salts
(c) Nitrosamine	 CI 37015	Special form of stabilized diazo
(d) Diazoamino		Special form of stabilized diazo
(e) Coupling component	 CI 37505 CI Azoic Coupling Component 2 Naphthol AS	Condensation products of 3-hydroxy-2-naphthoic acid with amines

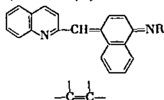
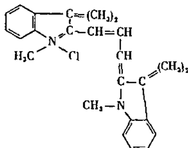
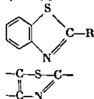
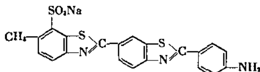
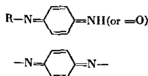
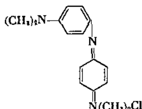
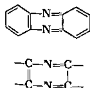
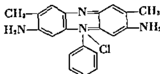
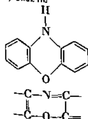
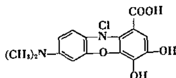
Some important dyes (Cont.)

Class basic chemical structure chromophore	Example	
(5) Stilbene 	Not known CI 40003 CI Direct Orange 15 Diamine Orange D	Condensation products of 5 nitro-*o*-toluene sulfonic acid with alkali forming first dinitrostilbene disulfonate which condenses *with itself or with hy*droxy and amino compounds
(6) Diphenylmethane (ketone imine) 	 CI 41000 CI Basic Yellow 2 Auramine 0	From Michler's ketone heated with ammonium and zinc chlorides
(7) Triarylmethane 		Condensation products of aromatic aldehydes with arylamines or phenols
(a) Diamino	 CI 42000 CI Basic Green 4 Malachite Green	
(b) Triamino	 CI 42500 CI Basic Red 9 Pararosaniline	
(c) Aminohydroxy	 CI 43520 Resorcine Violet	
(d) Hydroxy	 CI 43800 Aurine	

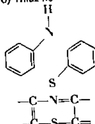
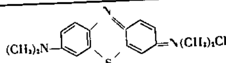
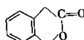
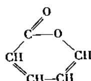
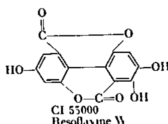
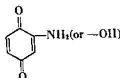
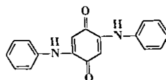
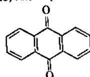
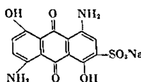
Some important dyes (Cont)

Class basic chemical structure chromophore	Example	
(e) D phenyl naphthyl methone	 CI 44045 CI Basic Blue 26 Victoria Blue B	
(8) Xanthene  		Basic mordant dyes subclasses are pyronines succineins sacchareins rosamines rhodamines rhodols fluorones (hydroxy and anthrahydroxy phthaleins)
(a) Am no	 CI 45170 CI Basic Violet 10 Rhodamine B	
(b) Hydroxy	 CI 45350 CI Acid Yellow 73 Fluorescein	Condensation of phthalic anhydride with hydroxy compounds
(9) Acrid ne  	 CI 46000 Acriflavine	Condensation of *m*-diamine with aldehyde cyclize and oxidize
(10) Qu nol ne  	 CI 47005 CI Acid Yellow 3 Quinoline Yellow	Condensation of quinolines with phthalic anhydride solvent and basic dyes for paper and wool when sulfonated give acid wool dyes

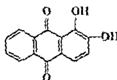
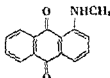
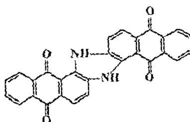
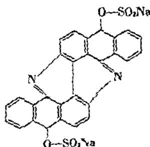
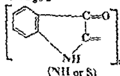
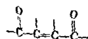
Some important dyes (Cont.)

Class, basic chemical structure, chromophore	Example	
(11) Methine and polymethine 	 CI 48070 CI Basic Red 12 Astraphloxine FF	Quinoline, benzothiazole, or trimethyl indoline nuclei linked together with methine chains; main uses in photography
(12) Thiazole 	 CI 49000 CI Direct Yellow 59 Primuline	Heat p-toluidine with sulfur, then sulfonate This structure is valuable if incorporated into other classes of dyes enhancing substantivity
(13) Indamine and indophenol 	 CI 49405 Bindschedler's Green	Oxidation of a p-diamine or p-aminophenol in presence of amine or phenol No use in textile dyeing; intermediates for sulfur colors; used in color photography
(14) Azine 	 CI 50240 CI Basic Red 2 Safranine T	Basic dyes; subclasses are quinoxalines, eurhodines and eurhodols, aposafranines, safranines, indulines, nigrosines
(15) Oxazine 	 CI 51030 CI Mordant Blue 10 *Gallocyanine*	Basic dyes; subclasses are monooxazines, dioxazines, oxazones

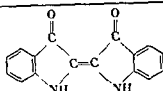
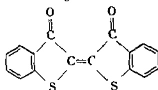
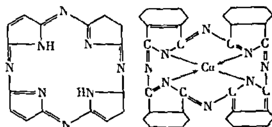
Some important dyes (Cont.)

Class, basic chemical structure, chromophore	Example	
(16) Thiazine 	 CI 52015 CI Basic Blue 9 Methylene Blue	Basic dyes
(17) Sulfur Structure not known; contains thiazole, thiazin, thianthrene rings with mercapto and polysulfide links; chromophore not known	CI 53185 CI Sulfur Black 1 Sulfur Black	Made by heating a variety of organic substances with sulfur and alkali polysulfides
(18) Lactone  	 CI 55000 Resoflavine W	Oxidation of polyhydroxy aromatic compounds; for chrome mordanted wool; of little importance
(19) Amino ketone and hydroxy ketone 	 CI 56000 Helindone Yellow CA	Of little importance
(20) Anthraquinone  		
(a) Acid	 CI 63000 CI Acid Blue 43 Alizarin Sapphire SE	Important wool dyes; sulfonated amino- or hydroxyanthraquinones

Some important dyes (Cont.)

Class basic chemical structure chromophore	Example	
(b) Mordant	 CI 58000 CI Mordant Red 11 Alizarine	
(c) Disperse	 CI 60505 CI Disperse Red 9	Containing no water solubilizing groups for acetate silk and other synthetic fibers
(d) Vat	 CI 69800 CI Vat Blue 4 Indanthrone	Most important class of dyes
(e) Esters of leuco vat dyes	 CI 70601 CI Solubilized Vat Yellow 1 Solubilized Flavanthrone	Water-soluble forms of vat dyes
(21) Indigoid  (NH or S) 		Vat dyes solubilized forms may also be made as in (20e)

Some important dyes (Cont)

Class basic chemical structure chromophore	Example	
(a) Ind goid	 CI 73000 CI Vat Blue 1 Indigo	
(b) Thioind goid	 CI 73300 CI Vat Red 41 Thioindigo Red B	
(22) Phthalocyanine	 CI 74160 CI Pigment Blue 15 Phthalocyanine Blue	Of great importance in pigments

natural fibers is often less satisfactory or impossible so that special classes of dyestuffs have been created for coloring these newer polymeric materials

Paper This is colored both by dyeing and by pigmentation with finely divided colored materials The color is usually added to the raw stock in the beater before the sheets are formed or on the calender Dyes are also applied in the coatings used on paper, such as the wax dye coating used for duplicating (carbon) paper *See* Paper and paper products

Leather One of the earliest materials to be colored leather has retained the use of natural dyes to a greater extent than most other materials but there are many synthet[illegible]

See Leather and fur processing

Wood Dyeing or tinting of wood is done with dyes in water, alcohol, or other solvents which evaporate after the wood has been painted with or soaked in the dye solution *See* Varnish

Pigments Many soluble dyes are converted into pigments by forming insoluble salts (that is replacing sodium in a dye salt with calcium) for use in lacquers, paints, and printing inks This insoluble salt, called a toner, may also be deposited on inorganic fillers such as aluminum hydroxide to form a lake

Food The appearance of many foods is enhanced by artificial coloring Butter and margarine are colored yellow, fruits and sauces which are often dulled by the process of preserving are brightened by addition of dye and soft drinks and candies are large consumers of dyestuffs In most countries only certain colors are permitted for this use In the United States all food coloring must be done by a natural dye or by choosing a synthetic from a list of certified food colors, of which the harmlessness to the human system is under constant check and review by a federal agency Elaborate studies of the toxicity and pharmacology of any new color must be provided by the manufacturer before it is admitted to the list

Oils Lubricating oils and gasolines are colored for improvement in appearance and for identification of grade, type of use, or merely as a trade mark of the manufacturer Waxes, shoe polishes and candles are also colored

Plastics and rubbers Coloration of resins plastics and elastomers may be done by solution of a dyestuff or by dispersion of a pigment in the sub

strate In this coloring operation there is no need for substantivity and choice of color is determined solely by the solubility shade and fastness properties desired

Biological samples The dyeing or staining of tissues of animal and vegetable matter is an important technique in physiological and medical studies Microorganisms and cell structure are made more visible and differentiated under the microscope by selective staining with dyes of varying affinity for protoplasm *See* STAIN (MICROBIOLOGICAL)

Photography Certain dyes (cyanines of the polymethine class) are added to photographic emulsions to increase sensitivity to light in special regions of the spectrum Color photographs are produced by formation of dyes from their components within the emulsion layer *See* PHOTOGRAPHY COLOR

Indicators Some property of the environment may bring about a change in hue in certain dyes thus giving information to the observer about this property Indicators usually respond to acidity alkalinity and oxidation or reduction but may respond to heat humidity electricity and water hardness *See* INDICATOR ACID BASE

Miscellaneous Soap synthetic detergents cosmetics ink hair fur metals anodized aluminum and many other materials are colored with dyes of various types Dyes are also used to produce colored smokes particularly for military identification *See* INK SMOKE

Methods of application This classification is used most frequently by the practical dyer

Acid dyes These are salts of organic acids (sulfonic and carboxylic) and are usually marketed as the sodium salts The acid groups confer water solubility on the dyestuff molecule When dissolved the dye ionizes (separates into particles with opposite electric charges) with the dye structure in the anionic (carrying the negative charge) part These dyes are used principally for wool natural silk synthetic fibers of polyamide and polyacrylic nature leather and paper They are normally applied to the fiber in a solution containing some sulfuric or acetic acid although some acid dyes will be fixed on the fiber from a neutral bath

Because of improved fastness resulting from treatment of the dyed material with metal salts especially chromium a large number of acid dyes are available which contain metal atoms as part of the anion as distinguished from the cationic metal atoms which form the salts of the organic acid These anionically bound metal atoms do not exhibit the usual reactions of metal ions and are said to be chelated These metallized (or premetallized) dyes have markedly superior fastness approaching the vat dyes in this property

Some organic chemical compounds (mainly stilbene derivatives) have substantivity for fibers but do not absorb light in the visible spectrum and hence show no color They do transform some of the ultraviolet light which they absorb into visible light thereby increasing the amount of white light reflected from them This gives a bluing effect to yellowed materials and makes them appear whiter These products are used on cloth and paper and are called white dyes optical bleaches or optical brighteners

Basic dyes These too are salts but of organic bases containing amino and imino groups The colored base is combined with a colorless acid such as hydrochloric or sulfuric and in solution the dye structure is in the cation (carrying the positive charge) These dyes have exceptional brightness but generally have only fair to poor fastness except when applied to acrylic fibers They have wide utility for coloring wool silk leather acrylic fibers and paper They can also be applied to cotton if the cotton has been mordanted (treated with tannin tannin and antimony salts and alum)

Mordant dyes Dyes which have little or no affinity for certain substrates may yet be fixed thereon if a mordant has first been applied The fixation of the color is principally the result of reaction with the mordant material These colors are also called adjective colors as compared to the direct dyeing substantive colors Basic dyes which are applied to mordanted cotton are commonly still called basic colors so that the designation mordant practically covers only acid dye types and alizarin The most common mordants are chromium salts for chrome dyeing processes Often the chromium compound is applied during the dyeing along with the dye or it can be applied after the dyeing These are called respectively meta mono or auto chroming and after or top chroming operations These treatments may not only change the shade of the original dyeing (the self shade) but they also may improve fastness of the dyeing to both water and light *See* MORDANT

Direct dyes These dyes are normally sodium salts of sulfonic acids and the colored part of the molecule is the anion They differ from the acid dyes in that they are so substantive to cotton or other cellulosic fibers that they are fixed on the fiber from an aqueous solution with the assistance only of additions of common salt or sodium sulfate to the dyebath Some authorities limit the designation substantive to these direct colors because of *their outstanding affinity These colors are also of* importance in coloring paper leather and silk and have many miscellaneous uses Because of the ease of application they constitute the bulk of the package dyes used by the housewife This is an important class of dyestuffs exceeded in numbers only by the acid dyes and in quantity used (in the United States) only by vat dyes

Many direct dyeings can be improved in wet fastness by an after treatment such as with copper or chromium salts formaldehyde resins and cationic fixing agents It is also possible to modify dyeings made by dyes of certain structures by a further chemical treatment (diazotization of a free amino group in the dye molecule followed by coupling with a developer) of the dye on the fiber Dyes suitable for such treatments are called developed

These treatments may or may not change the self shade

Ingrain dyes These are dyes which are formed directly on the substrate by some type of chemical action The principal subclasses are the azoic dyes and the oxidation dyes In a few instances dyes of the phthalocyanine type are developed on the fiber by special treatments

Azoic dyes are water insoluble azo compounds which have been formed within the substrate by chemical reaction of the intermediate components Generally the dyeing operation proceeds by dipping the cloth into a solution of one of the components (a hydroxy or amino component) drying the cloth without rinsing and then treating it with a solution of the other component (usually a diazo component which must be kept cold to prevent decomposition hence these colors are often called ice colors) Because the product of the reaction which will then be deposited throughout the fiber is a water insoluble color it is actually a pigment instead of a dyestuff

These colors are used for cotton and rayon and in great quantity in printing processes

Oxidation dyes are produced directly in the fiber or other substrates by a chemical oxidation following the impregnation of the substrate with certain aromatic amines The final product is probably a pigment instead of a dyestuff These colors are used principally for the dyeing of hair and fur

Disperse dyes These colors were originally developed for use on the synthetic fiber cellulose acetate This fiber has little affinity for the older known classes of dyestuffs Some very slightly water soluble colored materials which are chemically basic in nature transfer to this fiber from a water suspension if the colors are extremely finely divided particles This results in a solution of the dye in the solid fiber Use of these dispersed colors has extended to many of the newer synthetic fibers developed after cellulose acetate

Vat dyes This class of colors is distinguished by the special method of application needed—a vatting operation wherein the water insoluble color is made soluble by a chemical reduction of the chromophore the ketonic =C=O group to the ≡C—OH group the leuco compound which in the presence of alkali forms the water soluble leuco salt ≡C—OAlk This vat solution which is often a different color from the original insoluble material has affinity for cotton dyeing it with the shade of the vat solution Upon oxidation with air or oxidizing agents the reaction reverses to form the original water insoluble color leaving it deposited in the fiber as a pigment

One [illegible] trouble and expense of the vatting operation he need only apply to the cloth and then acidify and oxidize

The two major subclasses of vat colors are the indigos and the anthraquinones These latter colors are outstanding in their fastness to water, to light and to chemicals Since discovery of the first of the anthraquinones in 1901, this class has become the most important in dyestuffs

Sulfur dyes Dyes of this class are also applied by a vatting technique which makes the solubilized color substantive to cellulosic fibers These colors are made by treating a wide variety of organic compounds with sulfur and sodium sulfides With a few exceptions the final reaction products are not well identified chemical compounds with known structures These water insoluble dyes are dissolved in alkaline sodium sulfide solution which serves both as reducing agent and as source of alkali After application in the soluble leuco form oxidation produces the insoluble dye on the fiber

These colors are mostly used on cotton and viscose rayon They have moderate all round fastness and are relatively cheap Hence they are used in large quantities

Solvent dyes These colors are soluble in organic solvents such as benzene gasoline alcohol acetone oils fats and waxes Solvent dyes color merely by solution of the dye in the substrate They may be subclassified as spirit soluble colors with principal solubility in alcohol and oil soluble colors which are soluble in benzene and vegetable and mineral oils Uses include wood stains and varnishes lacquers printing and writing inks butter and margarine and plastics and resins

Fiber reactive dyes These are colors which have chemically reactive groups in their structure which can react with the substrate and thereby fix themselves by a conventional chemical union Such a process usually calls for special conditions and operations not common to dyeing procedure and must be justified by extraordinary fastness properties Introduction of these colors into commercial practice about 1950 has not resulted in any great volume of sales *See* HETEROCYCLIC COMPOUNDS QUINONE [W M]

Bibliography L F Fieser and M Fieser (ed) *Organic and Biological Chemistry* vols 2 and 3 1952 H A Lubs (ed), *The Chemistry of Synthetic Dyes and Pigments* 1955 U S Tariff Commission Rept *Synthetic Organic Chemicals* Ser 2 no 200 1957

Dyeing

The application of color producing agents to material usually fibrous or film in order to impart a degree of color permanence demanded by the projected end use True dyeing covers mechanisms in which molecules of material to be dyed become involved by various means with the molecules of the coloring matter or small aggregates thereof There is some overlapping between true dyeing and other methods of coloring which are called dyeing in the industry Products which are commonly dyed include textile fibers plastic films anodized aluminum fur wood paper leather and some food stuff

The broad term affinity is used to describe the various types of attraction between the material to be colored and the dye Affinity may be caused by

attraction between charged dye particles and oppositely charged dye sites on the material by various types of chemical attraction and by formation of solid solutions of dye in the material

When affinity is involved dyeing is an exothermic process which may be simply stated as follows

Dye in solution + undyed material →
dyed material + heat

This simplified equation explains the universally known fact that dye has a greater tendency to bleed into hot water than into cold water It also points out that more dye will ultimately be absorbed by a fiber or film at low temperatures than at higher temperatures when equilibrium is finally reached However the rate of dyeing increases geometrically whereas maximum dye absorption decreases only arithmetically with rise in temperature Maximum dye absorption is rarely necessary nor desirable therefore the trend in modern practice is toward dyeing at high temperatures and short times

Dyeing is accomplished by dissolving or dispersing the colorant in a suitable vehicle (usually water) and bringing this system into contact with the material to be dyed Although many dye molecules (or aggregates) may adhere to the material surface when they meet dyeing does not occur until the adhering dye particles migrate within the fibers or films All dyeing processes are designed to accomplish ultimately penetration of the undyed substance by the colorant

Dyeing assistants These are materials which do not impart color to the product to be dyed but as the name suggests promote or retard dyeing Usually but not always they affect the dye molecule

Swelling agents are assistants which open up the structure of the fiber temporarily so that dye molecules or aggregates may enter more freely and come in contact with otherwise inaccessible dye-sites

Carriers are agents (often solvents of low water solubility) which accelerate dyeing by breaking up or dissolving dye aggregates and bringing them to the fiber water interface in a size small enough to be absorbed by the material Frequently a carrier may exert a swelling action on the fiber in addition to its normal function

Dye retarders are a class of dyeing assistants usually inorganic or organic salts which slow up the dyeing process by forming evanescent compounds with the dye by buffering or depressing the ionization of an acid assistant or by temporarily occupying the more active or more accessible dye-sites on the fiber later to be dislodged therefrom by the dye

Aftertreating agents are salts resins or other products (more frequently applied to cellulosic fibers) to render the colored fabric more resistant to the effects of washing, perspiration or fading by ozone or combustion gases More often than not their application causes a loss in light fastness of the dyed material Aftertreatment with copper salts, for example is normally in order to increase light fastness whereas application of formaldehyde urea or melamine resins to increase wet fastness normally affects sunlight resistance adversely

Dyeing cellulose fibers Cotton and rayon are most commonly dyed by immersion of the fibers in a solution of direct dyes using an electrolyte such as common salt as assistant and then boiling this dyebath The affinity in this case may be a result of hydrogen bonding of areas in the dye molecule to hydroxyl groups in the cellulose more simply stated as adsorption Such dyeings usually exhibit only commercial (*minimum*) resistance to washing Treatment of the properly dyed fibers with resins and copper for example increases the resistance to washing with minimum loss of light resistance Some of the direct dyes have chemical groups (amino) which enable them to be converted after dyeing to large more-insoluble molecules by treatment with nitrous acid (diazotization) and a phenol or an amine (developing) Such dyeings are quite resistant to degradation by washing, but are usually characterized by poor resistance to fading by sunlight Naphthol dyeing is a special form of developed dyeing wherein the naphthol is a selected type of substantive developer and is applied to the fiber first Then a so-called diazo salt (not a dye) is added and an insoluble pigment forms within the fibers

Cellulose fibers may be dyed by vat dyes These dyes are normally insoluble pigments However under the influence of alkali and reducing agents they become water soluble and exhibit affinity for various fibers After dyeing the vat dyes revert to their insoluble form upon exposure to air or to oxygen supplying chemicals Indigo is a vat dye which does not possess affinity Its alkaline-reduced solution therefore can be impregnated into the fiber only by successive dips and oxidations Most vat dyes when properly applied are extremely resistant to degradation by such agents as washing bleaching sunlight and perspiration

Cellulose may be dyed by so called fiber reactive dyes which form tenacious bonds to cellulose and actually convert the cellulose to a different compound The connecting links between dye and fiber are considered to be ether or ester linkages

Sulfur dyes are applied to cellulose from a sodium sulfide solution and are characterized by subdued hues good wash fastness and low bleach resistance They are widely employed for the production of deep shades on cottons destined for use in work and play clothes

Basic dyes are dyed upon cellulose which has previously been mordanted with synthetic tannins or tannic acid and tartar emetic Dyeing is accomplished by the formation of a "lake" or pigment of the dye and mordant the function of the cellulose is merely that of a substrate in this type of dyeing Such dyeings are brilliant in hue but low in light resistance and for the latter reason the use of basic dyes on cotton is becoming obsolete

Dyeing animal fibers The dyeing of wool silk, and fur (felt) involves the formation of salts by reactive positively charged groups on the fiber and negatively charged color groups on anionic dyes

Level dyeing (acid or anionic) dyes which require a relatively large amount of strong acid to force the dye onto the fiber are used for carpet yarns felt hats and wherever even dyeing and penetration instead of washing fastness is a paramount requirement Milling dyes which demand less acid are faster to washing and are used for blankets and sweaters Where a very high resistance to light and washing is required acid dyeing or neutral dyeing metalized dyes are employed *See* DYE

Chrome (or mordant) dyes are applied to wool where subdued hues of excellent fastness are needed In the metachrome method chrome (sodium bichromate) is added to the dyebath at the start of the dyeing process If the wool is treated with the chrome mordant before dyeing the term chrome bottom is used and if the chrome is added at the end of the dyeing cycle the term top chrome or after chrome dyeing is applied If silk effect threads (such as in pin striped cloth) are present the after chroming is performed in a separate fresh bath a technique called the silk white process

Vat dyes and vat esters are also dyed on wool when bright shades of extreme fastness are required The leuco vat esters are developed by acid and oxidizing agents and the vat dyes are dyed as on cellulose but at lower temperature lower alkalinity and with a greater amount of reducing agent (sodium hydrosulfite) Fiber reactive dyes for wool have been developed

Dyeing man made fibers Cellulose acetate is usually dyed

suspens

fate)

lutions in cellulose acetate by passing from the water phase to the water fiber interface and migrating from the fiber surface inward In a technique similar to the diazotizing and developing technique described under cellulose dyeing water insoluble dye precursors are applied to the cellulose acetate diazotization is accomplished with sodium nitrite and acid β hydroxynaphthoic acid or hydroxanil are added to develop an insoluble colorant within the fiber Cellulose acetate may also be dyed by immersion in alcoholic water solutions or in formic acid water solutions of certain water soluble acid dyes

Polyamide (nylon) fibers are dyed by methods similar to those of wool dyeing with acid metalized acid neutral metalized fiber reactive and mordant dyes Vat dyes are applied by methods used for cotton dyeing but higher temperature and swelling agents such as o phenylphenol are used

washing and light is less important

Acrylic fibers such as Orlon and Acrilan 16 are dyed in light shades with disperse (acetate) dyes at high temperatures Being negatively charged the acrylics are most successfully dyed to heavy hues with cationic (positively charged) colorants By simultaneous use of copper salts and reducing salts acrylics may be dyed with anionic dyes (the so-called cuprous ion method) The positive cuprous ion forms a link between the negative dye molecule and the negative fiber molecule

Modified acrylics (Acrilan Zefran Creslan and Verel for example) are dyed with disperse and basic as well as with dyes normally applied to natural fibers The modification of the fibers involves the introduction of basic materials which form dye sites for anionic dyes

Polyester fibers (Dacron Terylene and Kodel) are dyed with disperse dyes at high temperatures or by the

as salicyli

bers may

peratures by leuco vat esters and disperse dyes or by applying these same dyes drying, and passing through a hot flue or over heated rollers at about 400°F (Thermosol method)

Polyethylene fibers may be dyed by using selected solvent soluble dyes in a mixture of ethylene glycol water and toluene

P lored by an

ness of pale shades to most normal color destroying effects and of heavy shades to all but rubbing-off is characteristic of fabrics colored by pigment dyeing however the bonding agent sometimes stiffens sheer goods

Dope or spin dyeing Melt spun fibers such as nylon Dynel or Dacron may be colored in the mass The colored plastic is melted and then extruded through spinnerets (refined shower head) and the congealed streams are collected and processed to yarn quality

Fibers such as cellulose acetate acrylics and viscose rayon are prepared by extruding their solutions into hot air or a coagulating water bath These so-called dopes are colored with pigments before extrusion and the coagulated fibers collected and processed into yarn

Textile dyeing equipment There are two fundamental types of dyeing processes—the dye liquor is pumped through stationary material or the material to be dyed moves through the dye liquor

Kiers are tanks which hold loose fiber or packaged yarn and the dye liquor is pumped through the fiber at temperatures up to the boiling point or (if the fiber can withstand it) under pressure at temperatures up to 270°F In one machine, goods are wound upon a perforated roller (or beam) and the whole placed in a pressure chamber Dye liquor under pressure above 212°F is pumped through the goods Several other machines which function by pumping dye liquor through stationary material are used in modern mills

In continuous dyeing cloth is impregnated with dye and then passed through a series of developing washing and drying zones to a final takeup roll One machine uses a passage through molten Wood's metal to develop the dye saturated goods whereas another uses hot oil for this purpose

The pad steam system consists of rollers which apply dye and necessary chemicals to the goods development takes place on continuous passage through a steam chamber

The pad roll system uses an insulated oven in which a huge roll of goods previously saturated with dye solution slowly turns until the dye has become fixed to the fiber

A jig promotes dyeing by winding goods through the dye bath back and forth from one roll to an other

A beck consists of an elliptical reel which draws goods *which have been sewn in an endless chain* from the dye bath and plaits it back into the bath repeatedly until dyeing is complete

Nontextile materials Dyeing anodized alumi num was formerly believed to be accomplished by the formation of aluminum salts of dyes by the electrically deposited aluminum oxide coating However many dyes having no salt forming groups were found to dye this film and it is postulated that at least in these cases conical craters receive the colorant and subsequent sealing by boiling in water or in solutions of salts in water converts the oxide to larger hydrated molecules (for example) and imprisons the dye aggregates in the sealed off cones At any rate anodized aluminum is readily dyed by many textile dyes Light and washing re sistance undreamed of in textile applications of some of these same dyes is achieved

Paper Paper pulp is usually dyed in the paper beater by dyes normally employed for cotton on occasion it is tinted by wool dyes and it is fre quently tinted by addition of pigments to the beater Finished paper is also colored by passing it over rollers which supply dye or colored coatings to its surface (calender staining)

Wood Wood is normally stained with solutions of dyes or dispersions of pigments in water in sol vents or in lacquers Wood however is frequently dyed by the application of water solutions of dyes at high temperature under pressures of 100 psi or more Freshly felled trees have been dyed by force pumping dye solution through the length of the log The dye solution replaces the natural moisture and sap An inflated tubular gasket seals the log end to the pressure cylinder in this method Wood destined to be cemented to corks for liquor bottle closures has been dyed with sulfur dyes to ensure resistance to alcohol

Leather Leather is dyed at low temperatures with the classes of dyes normally used for wool and cotton excepting vats and naphthols Formic acid is normally used to exhaust the dye For dress gloves leather is usually colored by applying the dye on the grain surface leaving the flesh side un dyed Leather is also dyed with natural dyes such as logwood fustic and quercitron

Fo [illegible]

[illegible]

[illegible]

ood dyes then washed and placed in flavored syrup Glazed citron citrus or watermelon rinds are colored by long immersion followed by drain ings in a series of food dye-colored increasingly strong sugar syrups Pistachio nuts in the shell are dyed by applying food dye solutions with or with out salt during roasting operations The dyeing of Easter eggs is of course well known

Fur Fur is usually colored by impregnations with synthetic organic intermediate compounds (amines) which are then oxidized to colored pig ments (azines) with peroxides for example When not fully developed some of the products produce objectionable physiological reactions so that great *care is taken to assure the absence of unreacted* amine in the dyed fur

Human hair It is not generally realized that hu man hair is in the final analysis a class of textile fiber Cosmetic companies appreciating this fact have hired textile experts to do research on dyeing waving and washing hair The dyeing of hair is of course limited to processes employing relatively low temperatures Food dyes (also carbon black) mixed with citric or malic acid are used for so called rinses Other rinses employing basic dyes and salt or textile dyes diluted with salts deter gents and malic citric or tartaric acids are on the market Products selected (as a rule) from among the safer fur dye intermediates are used for more permanent and intense coloring techniques these employ hydrogen peroxide for developing Natural dyes such as the age old henna and some metallic salts are still encountered in some hair dyeing establishments Rinses which are used to mask the yellowish tinge in gray or white hair as a rule use blue or violet food or textile dyes and frequently contain optical bleaches—substantially colorless compounds which receive invisible ultra violet light and reflect it to the eye as a blue or violet fluorescence

Plastic films or products Many plastic materials may be dyed by processes similar to those employed for textiles Nylon cellulose acetate polyethylene and polyester plastics are dyeable with dyes which color the same materials in yarn form Solvents *such as pyridine dioxane alcohol and trichloro*-benzene which swell the plastic may be used in wa ter solutions or emulsions in dyeing the above materials as well as polystyrenes methyl methac rylates and vinyls solvent soluble dyes are used in these techniques *See* FIBER MAN MADE FIBER NAT URAL PLASTICS FABRICATION TEXILE CHEMISTRY

[J E LO]

Bibliography L Diserens P Wengraf and H P Baumann *Chemical Technology of Dyeing and Printing* vol 2 1951 R S Horsfall and L G Lawrie *The Dyeing of Textile Fibers* 1927

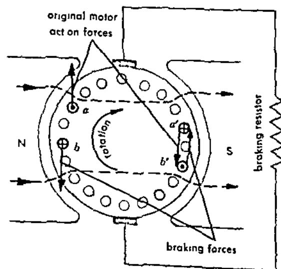
Dynamic braking

A system of braking a direct current motor by us ing the kinetic energy of the rotating elements to exert a retarding force The electric energy gen erated is dissipated in a resistor

The motor field remains energized b it the arma ture is disconnected from its voltage source and

connected across a resistor at the same time The motor becomes a generator with rotation due to the kinetic energy stored in the rotating elements The counter emf induced in the armature from generator action now causes a reversed armature current through the armature Forces due to this current act to retard the rotation such action diminishes proportionally with the speed of the armature

The schematic diagram shows typical conductors *a a* carrying current due to the applied armature voltage during normal operation of the motor The resultant motor action produces clockwise rotation When the armature voltage is removed the counter emf will force currents in the opposite direction through the conductors as shown by *b b* positions The distortion of the field around such conductors results in forces to oppose the original rotation and the motor armature is decelerated rapidly



Dynamic braking action

The braking resistor should be of proper wattage capacity to dissipate the armature current The resistance value in ohms affects the time of deceleration It is generally connected across the armature by means of contactors associated with controllers or relays controlled by push buttons

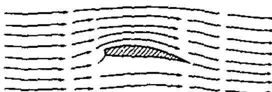
This method of motor braking may be applied to all types of dc motors used for control industrial or traction purposes *See* DIRECT CURRENT MOTOR

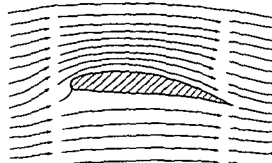
[L F C]

Bibliography C L Dawes *A Course in Electrical Engineering* vol 1 4th ed 1952 A E Fitzgerald and C Kingsley Jr *Electric Machinery* 1952

Dynamic similarity

A relationship existing between two homologous fluid flow systems such that the corresponding parts of the systems experience similar net forces Dynamically similar flows about geometrically similar bodies will themselves be geometrically similar


flow A


flow B

Dynamically similar flows about similar airfoil profiles.

as illustrated Consequently this concept is basic to the meaningful extrapolation of model results to full scale performance However geometrically similar flows are not necessarily dynamically similar

Dynamic similarity between two flow systems will occur if certain nondimensional parameters formed from the flow variables have the same values for both systems

One important parameter for establishing dynamically similar flows is pressure coefficient $p/\rho V^2$ where p ρ and V are respectively a reference pressure density and velocity Other important nondimensional parameters are given in the accompanying table along with associated physical effects which characterize the parameters

Dynamic similarity parameters

Nondimensional parameter*	Name	Physical effect
$\rho VL/\mu$	Reynolds number	Viscosity
V/c	Mach number	Compressibility
V^2/Lg	Froude number	Gravity
x_m/L	Knudsen number	Pressure
$\rho V^2L/\sigma$	Weber number	Surface tension
$c_p\mu/\kappa$	Prandtl number	Heat conduction
$\beta TgL^3\rho^2/\mu^2$	Grashof number	Free convection

* The reference variables are defined as follows L length μ coefficient of viscosity c speed of sound g acceleration of gravity σ surface tension κ coefficient of thermal conductivity β coefficient of thermal expansion and T temperature x_m mean free path of molecule c_p specific heat at constant pressure

The parameters in the table can be determined analytically by dimensional analysis or by examining the invariance of the differential equations and boundary conditions for the flow systems under scalar transformations of length time and mass Other useful parameters can also be defined Often

these can be obtained as ratios of the tabulated parameters

In practice, it is often difficult to establish equality of all similarity parameters simultaneously for two flows Equality of parameters corresponding to dominant flow properties is usually sufficient Given low speed viscous flows for example the Reynolds numbers of the two flows would be equated but the Mach numbers might be ignored See DIMENSIONAL ANALYSIS, FLUID MECHANICS, FROUDE NUMBER, MACH NUMBER, MODEL THEORY

[A G H A]

Bibliography: H L Langhaar *Dimensional Analysis and Theory of Models*, 1951 S Pai *Viscous Flow Theory*, vol 1, 1956, A F Zahm *Theories of Flow Similitude*, NACA TR 287 1928

Dynamical analogies

Analogies are useful when one is comparing a familiar system with an unfamiliar one An electrical circuit can be considered to be a vibrating system and this immediately suggests analogies between electrical circuits and vibrating systems The work of an acoustical engineer involves the study of acoustical, electroacoustical mechanoacoustical, or electromechanoacoustical systems while a mechanical engineer studies vibrating systems involving masses springs and friction The analogies between systems of these types and electrical circuits are known as dynamical analogies

Vibration problems may be solved by establishing an equivalent electrical circuit This is known as an electromechanical analogy method By this experimental technique it is possible to study the effect of varying certain parts of a mechanical system such as damping or spring rate The corresponding electrical components are easily controlled and are inexpensive to provide The electrical equivalent system will not only save time but may be the only means to solve some complex mechanical problems

The analogy is based on a similarity of the equations of electrical circuits with those of the mechanical system The circuit equations are generally established by a method based on Kirchhoff s second law which states that in any network the algebraic sum of the potential difference around any closed circuit is zero (*see* KIRCHHOFF S LAWS OF ELECTRIC CIRCUITS) Such a system is illustrated in Fig 1 An examination of this system shows it to be similar to a mechanical system for a forced vibration in a single degree of freedom system which is illustrated in Fig 2 The mass m in the mechanical system is equivalent to the inductance L, the damping factor c to the resistance R, and the spring constant k to the capacitance C The forcing function $F_0 \cos \omega t$ is analogous to the impressed voltage $v_0 \cos \omega t$

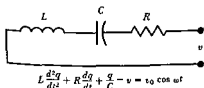
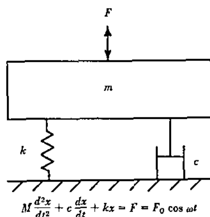


Fig 1 Electrical circuit



$$M\frac{d^2x}{dt^2} + c\frac{dx}{dt} + kx = F = F_0 \cos \omega t$$

Fig 2 Mechanical system

It should be noted in the mechanical system that the forces acting on the mass are in parallel while in the electrical system the components of the circuit are in series If the forces in the mechanical system were in series then the equivalent electrical system would be put in parallel The equivalents for the mechanical system and the electrical system are summarized here

Mechanical quantity	*Electrical quantity*
m mass	L inductance
k spring constant	$\frac{1}{C}$ $\frac{1}{\text{capacitance}}$
c damping factor	R resistance
x displacement	q charge
dx/dt velocity	dq/dt current
F force	v voltage
ω frequency	ω frequency

See ALTERNATING CURRENT CIRCUIT THEORY, VIBRATION DAMPING [K W J]

Bibliography H F Olson, *Dynamical Analogies*, 1958

Dynamics

That branch of mechanics which deals with the motion of a system of material particles under the influence of forces, especially those which originate outside of the system under consideration From Newton's third law of motion namely, to every action there is an equal and opposite reaction, the internal forces cancel in pairs and do not contribute to the motion of the system as a whole, although they determine the relative motion if any, of the several parts

Particle dynamics refers to the motion of a single particle under the influence of external forces, particularly electromagnetic and gravitational forces The dynamics of a rigid body is the study of the motion under given forces, of a system of

particles the distances between which are postulated to be constant throughout the motion

In classical dynamics the basic relation that enables the motion to be determined once the force is known is Newton s second law of motion which states that the resultant force on a particle is equal to the product of the mass of the particle times its acceleration For a many particle system it becomes impracticable to write and solve this equation for each individual particle and in general the motion may be computed only on a statistical basis (that is by the methods of statistical mechanics) unless as for a few particles or a rigid body the number of degrees of freedom is sufficiently small *See* Degree of freedom (mechanics) Kinetics (classical mechanics) Newton s laws of motion Rigid body dynamics Statistical mechanics *see also* Kinematics [H C C]

Dynamo

An electric machine for the conversion of electrical energy into mechanical energy or conversely mechanical energy into electrical energy It is called a generator if it converts mechanical into electrical energy and it is called a motor if it converts electrical into mechanical energy *See* Electric rotating machinery Generator electric Motor electric [A R E]

Dynamometer

A special type of electric rotating machine used to measure the output torque or driving torque of rotating machinery Most dynamometers consist of a direct current (dc) machine with the stator cradle mounted in antifriction bearings The rotor is connected to the rotor of the machine under test The field current is introduced through flexible leads The stator is constrained from rotating by a radial arm of known length to which is attached a scale for measuring the force required to prevent rotation The torque of the connected machine is found from the product of the lever arm length and the scale reading after correcting the scale reading by the amount of the zero torque reading By using a tachometer to measure the rotor speed the power may be found from the equation hp = $2\pi NT/33\,000$ where N is the shaft speed in rpm and T is the torque in foot pounds

If the machine under test is a motor the dynamometer will act as a generator The dynamometer output is absorbed by a loading resistance or by feeding it into a dc line The amount of the output is easily adjusted by changing the loading resistance or by changing the field excitation If the machine under test is a generator or mechanical load the dynamometer will act as a motor The speed is adjusted by changing the armature voltage or the field excitation The dynamometer method is direct reading and is more accurate than measuring the electrical output and correcting for the losses Except for the inaccuracy caused by friction in the stator mounting bearings and by windage loss not reflected in the stator torque all of the shaft torque is accounted for in the scale reading

When the machine under test is a motor a mechanical device known as a prony brake may be employed to convert the output energy to heat through friction A drum which may be water cooled is driven by the machine under test A brake arm which is constrained from rotating by a scale at the outer end is tightened around the drum to increase the friction and to produce the desired torque The torque is the product of the scale reading and the length of the brake arm. This is an inexpensive and accurate method of measuring the torque of small motors With large horsepower motors however, it becomes difficult to dissipate the large amounts of energy Large capacity units which utilize a liquid brake in place of the friction drum have been constructed These are smaller and less expensive than electric units of like capacity but lack the flexibility and ease of recovering the energy [A R E]

Dyne

A unit of force in the centimeter gram second system of units The dyne is based upon a mass unit (gram) which is 1/1000 kilogram and a length unit (centimeter) 1/100 meter, hence 1 dyne is the force which imparts 1 cm/sec^2 acceleration to a 1 gram mass One dyne is 10^{-5} newton *See* Force [C E F]

Dyschondroplasia

A deforming nonhereditary disease of early youth also referred to as Ollier's disease or enchondromatosis It varies greatly in severity and may affect one or more bones The fingers and the long bones of the extremities are most commonly affected while skull ribs and pelvis are rarely involved The disease is progressive and involves the growing portions of the affected bone

In normal growth a bone increases in length by continuous production of new bone within the caps of growing cartilage (epiphyses) at its ends In dyschondroplasia the growth rate of the epiphyseal cartilage is retarded and formation of new bone is defective leaving irregular residual islands of abnormal cartilage scattered along the length of the shaft These two factors result in an abnormally short bone and irregular nodular swellings along the shaft of the bone The former results in arms or legs of unequal length, the latter may cause such irregular enlargement of fingers that function is severely impaired The deformities produced increase steadily until skeletal maturity is reached at which time the process slows down or stops entirely always leaving some degree of residual disability

No cure is known but operation designed to shorten the normal mate of an affected bone is sometimes resorted to in order to lessen the inequality in length *See* Skeletal system disorders [W R AD]

Dysprosium

Element number 66, dysprosium Dy is a metallic element belonging to the rare earth group Its atomic weight is 162 51 and the naturally occurring element is made up of the stable isotopes



Dy^{156} 0 0524% Dy^{158} 0 0902% Dy^{160} 2 294%, Dy^{161} 18 89% Dy^{162} 25 53% Dy^{163} 24 97% Dy^{164} 24 18% Dysprosium was discovered by L de Boisbaudran in 1886 It forms a white oxide, Dy_2O_3 which dissolves in acid to give a yellowish green solution For properties of the metal, *see* RARE EARTH ELEMENTS

The metal is attacked readily by air at high temperatures but at room temperatures in massive blocks is fairly stable in the atmosphere and remains shiny for long periods of time The metal is paramagnetic but as the temperature is lowered, it becomes first antiferromagnetic and then ferromagnetic The Neel point is about 178°K and the Curie point is about 85°K At very low temperatures the metal shows strong anisotropic magnetic properties It is easy to saturate the metal in the direction of the hexagonal planes but it is almost impossible with the fields available in the laboratory to do so at right angles to the plane [F H SP]



e—Enstatite

e

The number e is usually defined as the limit approached by the expression

$$\left(1+\frac{1}{n}\right)^n$$

as n approaches infinity. If the given expression is expanded by the binomial theorem and if one uses the theorem that the limit of the quotient of two polynomials of equal degree as the variable tends to infinity is equal to the ratio of the coefficients of highest degree, one obtains the expansion

$$e = 1 + \frac{1}{1} + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} + \cdots + \frac{1}{1 \cdot 2 \cdot 3 \cdots n} + \cdots$$

Clearly e is larger than 2; it may be easily shown that it is smaller than 3. It can also be shown by elementary methods that e is irrational, that is, it cannot be represented as the quotient of two integers. Furthermore, e is transcendental; it does not satisfy any algebraic equation with integral coefficients. The transcendality of e was proved by the French mathematician C. Hermite in 1873; the proof constitutes an important milestone in the history of mathematics.

By the method outlined above it may be shown that the limit of

$$\left(1+\frac{x}{n}\right)^n$$

as n tends to infinity is e^x, and moreover, that

$$e^x = 1 + \frac{x}{1} + \frac{x^2}{1 \cdot 2} + \cdots + \frac{x^n}{1 \cdot 2 \cdot 3 \cdots n} + \cdots$$

The function e^x is of great importance in mathematical analysis and is encountered in numerous problems in applied mathematics. For this reason it has been extensively tabulated. The most recent tables of exponentials for both positive and negative values of x were computed by the Computation Laboratory of the National Bureau of Standards. One of the most important formulas in mathematics involving e is Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$. This is readily obtained from the above expansion of e^x by replacing x by $i\theta$. The real part of the expansion is recognized as the expansion of $\cos\theta$ while the coefficient of i is recognized as the expansion of $\sin\theta$. An immediate consequence of Euler's formula is de Moivre's formula $(\cos\theta + i\sin\theta)^m = \cos m\theta + i\sin m\theta$. If the first member of the last equation is expanded by the binomial formula, and if the real and imaginary parts of both members are equated, one obtains two important formulas which permit the evaluation of $\cos m\theta$ and $\sin m\theta$ in terms of $\cos\theta$ and $\sin\theta$. *See* BINOMIAL THEOREM; CALCULUS, DIFFERENTIAL AND INTEGRAL; LOGARITHM. [A. N. L.]

Eagle

Any of several large members of the family Accipitridae differing from the hawks primarily in size. The genus *Aquila* has eight species found on all continents except Australia and South America. In this genus the entire tarsus is feathered. In the genus *Haliaeetus*, the lower one-third of the tarsus is unfeathered. There are also eight species in this genus, of cosmopolitan distribution except for South America. One member of each genus occurs in the United States. The bald eagle, *Haliaeetus leucocephalus*, the national symbol of the United States, is the better known of the two. It is still reasonably common in Alaska but is scarce elsewhere. The golden eagle, *Aquila chrysaetos*, is a Holarctic species found in the United States primarily from the Rocky Mountains westward, al-



The golden eagle, *Aquila chrysaetos*, length to 35 in. (From J. G. Wood, *Popular Natural History*, Porter and Coates, 1885)

though it is also present in the mountains of the Southeast and occasionally elsewhere See FAL CONIFORMES HAWK [J D B]

Ear

The vertebrate ear is an organ of hearing and of equilibration because it is concerned with auditory sensations and also with sensations of position and of rotation

In the mammal, this organ consists of three parts the external ear which receives the sound waves the middle ear which transmits the vibrations by a series of three small bones and the internal ear a complex bony chamber placed deep within the skull The internal ear contains a membranous labyrinth having localized areas of sensory epithelium one of which the organ of Corti is auditory in function whereas other sensory areas the cristae and maculae are equilibratory The membranous labyrinth is surrounded by and contains special aqueous fluids (Fig 1)

Sound waves are conducted by air in the external ear by solid structures in the middle ear and by aqueous media the perilymph and endolymph in the internal ear The internal ear is found in all vertebrate groups and is the essential part of the mechanism The middle ear is found in amphibians reptiles birds and mammals but the external ear is characteristic of mammals

Equilibration seems to have been the primitive function of the internal ear because in many fishes *amphibians* and reptiles the structure of the internal ear is quite uniform and most of its component parts are related to equilibration rather *than to hearing* There is in fact surprisingly little variation in the differentiated structures of the internal ear that affects equilibration through*out the* vertebrate series In reptiles and birds however and especially in mammals auditory structures become increasingly prominent until they tend to overshadow equilibratory structures *See* EQUILIBRIUM BIOLOGICAL HEARING

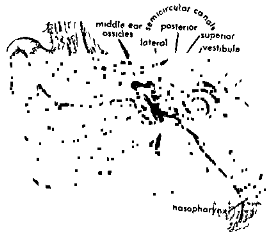


Fig 1 Schematic drawing of the human ear (Drawing by M Brodel Three Unpublished Drawings of the Anatomy of the Human Ear, Saunders)

THE EXTERNAL EAR

The external ear consists of the auricle or pinna and the external acoustic meatus which receive and conduct the sound waves to the tympanic membrane which closes the internal end of the meatus The external ear is first indicated in certain reptiles in which the tympanic membrane, instead of being at the surface, is sunk into the head with the formation of an external acoustic meatus The auricle may be indicated externally as folds of skin In birds, the meatus consists merely of a short bent tube that has small, stiff feathers associated with it

Acoustic meatus. In mammals, the external acoustic meatus is a canal which has cartilage and bone in its wall and which is lined with a cutaneous layer The meatus is longer and the tympanic membrane is farther removed from the surface than in reptiles and birds The auricle is a fold of skin above and behind the external opening of the meatus and contains an irregular cartilage continuous with that of the meatus There are numerous elastic fibers in this cartilage so that the meatus is permanently open although somewhat yielding

A primitive type of ear cartilage is found in the Australian spiny anteater, one of the lowest mammals Here the cartilage of the auricle is a flat plate that is continuous with the cartilage of the meatus The meatus consists of crescentic cartilages which partially encircle the meatus and which are united by a central longitudinal cartilaginous bar The great variations found in the shapes of the ear cartilages of higher mammals can be referred to separation, fusion, variations in size or complete absence of elements of this fundamental pattern

Musculature. The external ear has certain voluntary muscles In the human, three extrinsic muscles, the anterior, the superior, and the posterior radiate forward, upward, and backward on the skull from their attachment to the auricle but are usually inactive In the human there are also six small poorly developed intrinsic muscles which stretch from one part of the auricular cartilage to another and so cannot possibly have any function

Such muscles are also found in other mammals in which they are more complex and better developed They can nevertheless be homologized with those in the human, and as in the human they are innervated by the facial nerve Extensive movements thereby become possible, and the auricle can be turned backward, forward or laterally, as in the horse

THE MIDDLE EAR

The concept of the middle ear as a group of associated structures is derived from human and mammalian anatomy In these forms, the middle ear comprises the tympanic cavity and the auditory or eustachian tube The tympanic cavity is an air containing space in the temporal bone, bounded laterally by the tympanic membrane and medially

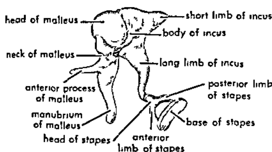


Fig 2 The auditory ossicles of the right middle ear (From J P Schaeffer, ed, *Morris Human Anatomy* 11th ed, McGraw Hill 1953)

by the petrous part of the temporal bone which contains the internal ear

Tympanic membrane. The mammalian tympanic membrane is a thin structure with an outer cutaneous layer and an inner mucous layer continuous with the lining of the tympanic cavity Between the two are two layers of collagenous fibers and fibroblasts In an outer layer, the fibers radiate from the attachment of the handle of the malleus in the inner layer the fibers are circular In a flaccid portion of the membrane radial fibers are lacking The membrane in mammals is circular oval or bean shaped and the length of the radial connective tissue fibers varies in a single membrane

Osteology. In mammals a chain of three small bones, the malleus incus, and stapes (hammer anvil and stirrup), extends from the tympanic membrane across the middle ear cavity to an oval opening in the temporal bone (Fig 2) The malleus is attached by its handle almost at the center of the tympanic membrane, the incus is intermediate and the base of the stapes fits into the oval window or fenestra vestibuli in the temporal bone Tiny articulations connect the bones This system of three bones acts as a bent lever to convert the vibrations of the tympanic membrane into thrusts intensified in force but decreased in amplitude of the stapes against the perilymph the fluid that surrounds the membranous labyrinth Sound waves are thus transmitted across the tympanic cavity

Musculature The mammalian tympanic cavity contains the tendons of two small muscles The tensor tympani muscle tendon is attached to the handle of the malleus to tighten the tympanic membrane and so protect it against excessive vibrations The stapedius is the smallest of all skeletal muscles and is inserted on the stapes It influences the tension of the chain of bones and prevents excessive pressure of the stapes against the perilymph

Tympanic cavity. The posterior part of the mammalian tympanic cavity is continuous with air filled cavities in the mastoid process of the temporal bone The tympanic cavity is also continuous with the nasopharynx through the auditory or eustachian tube whose wall is partly bony partly cartilaginous and partly fibrous For most of its length the walls of the auditory tube are normally in contact except during swallowing when contractions of neighboring muscles open the tube, thus permitting air pressure on the internal surface of the tympanic membrane to be equalized with outside pressures

In mammals the structures of the middle ear may be designated as sound conducting and it may seem logical to equate the presence of middle ear structures with hearing However, in other vertebrates such designation is uncertain because adequate information is not available concerning *their capacity for hearing In certain fishes* hearing has been demonstrated although there is no middle ear Other structures must transmit the sound waves

Membranous labyrinth In many fishes the membranous labyrinth is found in a part of the skull that is incompletely separated from the cranial cavity Thus the membranous labyrinth obviously receives pressure changes occurring in the cranial cavity To a lesser extent this is true of amphibians and of certain reptiles—snakes and lizards In these forms transmission of pressure changes from the cranial cavity to the membranous labyrinth can easily be accomplished because parts of the membranous labyrinth the endolymphatic duct and sac and parts of the perilymphatic sac project through large apertures into the cranial cavity In other reptiles (crocodilians) and in birds and mammals the openings which in lower forms transmit the endolymphatic and perilymphatic sacs are very narrow and the sound waves must be transmitted to the perilymph through the oval window A round window closed by a secondary tympanic membrane provides a yielding structure which bulges into the tympanic cavity as the perilymph is compressed The importance of the round window and its membrane is indicated by the fact that they are found in all reptiles birds and mammals A somewhat similar membrane may also be found in some amphibians but here it is found in the base of the skull whereas in higher forms the round window is in a lateral position In the higher mem

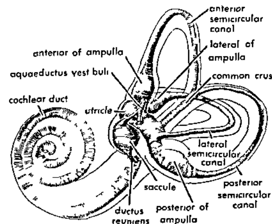


Fig 3 A medial view of the right membranous labyrinth in the human, projected on outline of the perilymphatic space (From W J S Krieg, *Functional Neuroanatomy*, 2d ed, McGraw-Hill 1953)

bers of the vertebrate series then the middle ear and its contained structures become more and more clearly related to hearing

Tympanic membrane In general the middle ear of Amphibia if such it may be called in the absence of an external ear resembles that found in the mammal A large tympanic membrane with a radial structure forms the lateral wall There is broad continuity between the tympanic cavity and the pharynx These middle ear structures are however found only in the tailless Amphibia In the tailed Amphibia and Apoda they are absent or greatly reduced

In reptiles and birds, with the exception of turtles and snakes there is in general a tympanic cavity that is continuous with the pharynx and closed externally by a tympanic membrane This space is not traversed by three middle ear bones as in mammals but by a rodlike columella which extends from the tympanic membrane to the oval window The columella may in some species be subdivided laterally into the extracolumella and medially into the stapes

In turtles the tympanic membrane is covered with scales and is so thick and rigid that it cannot possibly receive and transmit sound waves Snakes do not have a tympanic cavity there is however a columella which extends from the quadrate bone into the skull and comes into contact with the perilymph Birds have a tympanic cavity which extends into numerous air filled chambers of the skull In birds and crocodiles a muscle sometimes considered to be homologous with the stapedius muscle of mammals serves to prevent excessive excursion of the tympanic membrane

In reptiles birds, and mammals, a single structure the columella or stapes fills the oval window but in Amphibia there are characteristically two elements the plectrum or columella and the operculum The operculum is a part of the ear capsule that has become separate The plectrum extends across the tympanic cavity to the tympanic membrane if one is present, but whether the plectrum represents the columella of reptiles and birds is undecided The functional significance of many of these variable middle ear structures in the lower vertebrates is in need of further clarification

THE INTERNAL EAR

The internal ear of vertebrates is comprised of an intricate arrangement of structures The membranous labyrinth consists of a membranous chamber and of the semicircular canals connected with it The semicircular canals are filled with endolymph and surrounded by perilymph (Fig 3) A capsule of cartilage or bone, the bony labyrinth encloses each labyrinth and reproduces its structure in a simplified way An opening in the capsule transmits the nerves Connective tissue strands between the internal surface of the capsule and the membranous labyrinth hold the labyrinth in place

Semicircular canals. The three semicircular canals are found in the three planes of space and are attached to the utriculus, the dorsal portion of the membranous labyrinth Each semicircular canal has a slight enlargement the ampulla in the wall

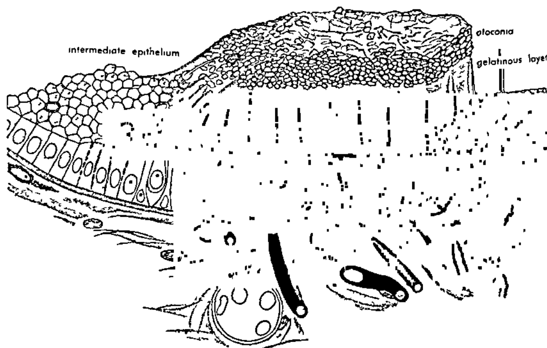


Fig 4 Plastic diagram of part of a macula The goblet shaped cells are sensory some are shown with nerve fibers surrounding them (After Kolmer in A A Maximow and W Bloom, Textbook of Histology 7th ed , Saunders, 1957)



Fig 5 Plastic dia part of a crista as seen in a longitudinal section of a canal and its ampulla The goblet-shaped cells are sensory some are shown with surrounding nerve fibers (After Kolmer in A A Maximow and W Bloom Textbook of Histology 7th ed Saunders 1957)

of which is found a sensory area the crista In the shark the semicircular canals are 18 5 cm long and in the whale 2 5 cm The sacculus is the ventral portion of the membranous labyrinth In its wall as in the floor of the utriculus is found another type of sensory area the macula It is to these and other sensory areas that the nerve fibers are distributed These structures of the internal ear are involved in equilibration the phylogenetically older function Hearing on the other hand is associated with other sensory areas found in a diverticulum of the sacculus

Membranous labyrinth The endolymphatic sac and its duct are other parts of the membranous labyrinth the functions of which are less clearly understood These structures arise very early in development as a medial diverticulum of the differentiating labyrinth In selachian fishes the duct communicates with the exterior In the frog the sac consisting of numerous communicating tubules filled with calcareous crystals is found in the cranial cavity and extends along the brain it is continued along the spinal cord and its lateral extensions surround the spinal ganglia In the gecko the sac extends beyond the cranial cavity into the neck and shoulder and into the orbit In mammals the endolymphatic sac which is small and lies within the cranial cavity external to the dura mater is apparently the site of elimination of endolymph

Histologically the membranous labyrinth consists of an epithelial layer and an external supporting connective tissue layer Except for the sensory areas the epithelium is a single layer of flattened cells Of particular functional significance are areas where the wall is thinner and where vibrations in the surrounding perilymph are probably transmitted to the endolymph In sensory areas the

epithelium is thickened and consist of sensory and supporting cells The sensory cells are tall and have sensory hairs which protrude into the laby rinth Sensory nerve fibers surround the bases of these cells

Sensory areas Three kinds of sensory areas may be distinguished on the basis of structure These are the maculae crista and organ of Corti or basi lar papilla

Macula The maculae have statoliths (otoliths) on their internal surfaces (Fig 4) these may be numerous small crystals in a gelatinous layer or large single stones as in fishes The macula of the utriculus forms part of the floor of the utriculus and is approximately horizontal The macula of the sacculus is part of the medial wall of the sacculus and is approximately vertical The maculae and their statoliths are presumably acted upon by grav ity and these sensory areas probably yield infor mation concerning the position of the head in space The macula of the lagena a knoblike out growth of the sacculus is more variable in struc ture and distribution It is found in a diverticulum of the sacculus In reptiles and birds it is associ ated with the basilar papilla or the organ of Corti Among mammals it is found only in monotremes Its function is unknown

Crista A second type of sensory area is the crista (Fig 5) a connective tissue ridge covered with sensory epithelium running transversely across the ampulla of the semicircular canal and constituting part of its wall A gelatinous structure the cupula is found on the surface of the epithelium and ex tends more than half way across the ampulla the long hairs of the sensory cells lie within this struc ture The cristae of the three semicircular canals one in each ampulla are supposedly stimulated by streaming movements of the endolymph Such rela tive movements occur in turning motions of the head because of the inertia of the fluid which tends to remain in place as the wall of the canal moves past it

The papilla neglecta is related structurally to the crista Its function is unknown but it is best developed in fishes It is absent in amphibians variable in reptiles and birds and absent in mam mals

Basilar papilla The third type of sensory area is the basilar papilla or organ of Corti associated with hearing (Fig 6) Typically a specialized membrane the tectorial membrane lies upon the sensory area but is attached at one side Such an organ is first found in the amphibian ear as the papilla amphibiorum Most amphibians also have a second sensory area or basilar papilla of this na ture which forms part of the wall of a diverticu lum of the lagena This sensory area evolves into the most highly differentiated structure of the ear the organ of Corti of the mammal

In reptiles the basilar papilla undergoes marked evolution It lies on a thin part of the membranous wall the basilar membrane which alone separates it from the perilymph In various reptilian species the diverticulum and its basilar papilla are elon gated most markedly so in the crocodile The elon gated diverticulum is known as the cochlear duct In birds the basilar membrane and the basilar papilla which it bears are stretched between car tilage walls which separate one perilymphatic space the scala vestibuli from another the scala tympani The cochlear duct thus lies between two perilymphatic channels as in mammals

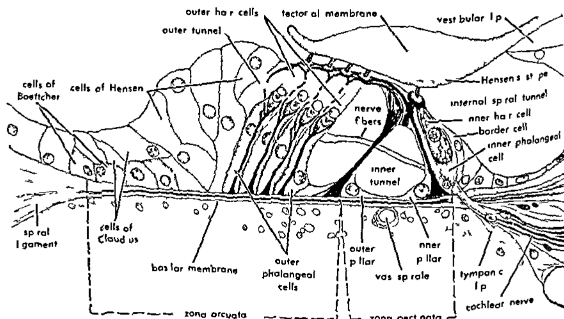


Fig 6 A radial section of the *human organ of Corti* (After Held in A A Maximow and W Bloom *Textbook of Histology* 7th ed Saunders 1957)

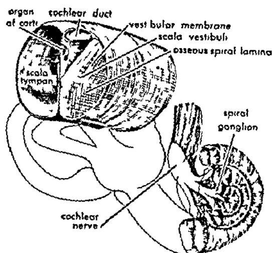


Fig 7 Section of the cochlea through the modiolus, showing spiral ganglion and cochlear nerve Above and left an enlarged segment of one turn of the cochlea (From W J S Krieg, Functional Neuroanatomy, 2d ed, McGraw Hill 1953)

The organ of Corti of mammals including man as in reptiles and birds, rests on a basilar membrane stretched between bony parts As in birds there are tall supporting phalangeal and pillar cells and sensory hair cells in this thickened epithelium the sensory cells being found in an upper layer, toward the interior of the cochlear duct Endolymph filled spaces are present in the epithelium (inner and outer tunnel, spaces of Nuel) The "inner" hair cells are placed in a single row the "outer" hair cells in three to five rows Each hair cell has numerous hairs extending upward In the lower mammals, there are 8–12 The human has the largest number of hairs, more than 100 per outer hair cell There are estimated to be 3500 inner hair cells and 12,000 outer hair cells in the human

The cochlear duct The cochlear duct of the mammal is coiled about a bony axis the modiolus, but the amount of spiralling varies In the monotremes the duct makes barely ½ turn, in the guinea pig 4–4½ turns The human cochlear duct (Fig 7) which is 35 mm long, makes 2½ turns Within the bony modiolus is the cochlear division of the acoustic nerve Its ganglion cells constitute the spiral ganglion so called because the cells parallel the coils of the cochlear duct within the modiolus The cells are bipolar, one process extending outward to the hair cells, the other to the brain

The mammalian cochlear duct is more or less triangular in cross section Its outer wall consists of a thick capillary containing epithelium, the stria vascularis, which is regarded as the site of continuous endolymph formation The tectorial membrane rests on the hair cells within the cochlear duct It is attached to tall supporting cells that line the inner part of the spiralling duct The rest of the duct wall is formed by a single layer of flattened cells, as in other nonsensory parts of the membranous labyrinth

Basilar membrane The basilar membrane extends from the modiolus to the bony side wall of the cochlea The number of its radial connective tissue fibers is estimated at 24 000 These fibers supposedly act as a resonating mechanism so that specific portions of the organ of Corti lying upon them may be stimulated by tones of different pitch This is possible because fiber length increases from base to apex of the cochlea, at the beginning of the first coil the fibers are 0 064–0 128 mm long, whereas at the apex they measure 0 35–0 48 mm The basilar membrane and the cochlear duct rest ing upon it end at the apex of the cochlea, so that the two perilymphatic spaces the scala vestibuli and scala tympani on opposite sides of the cochlear duct become continuous at the apex the so called helicotrema Thus a perilymph channel free of connective tissue extends from the oval window, where vibrations are received from the stapes along the scala vestibuli to the helicotrema and finally along the scala tympani to the round window and the secondary tympanic membrane that closes it

DEVELOPMENT

The membranous labyrinth is first indicated in early embryonic stages as an ingrowth of the surface layer ectoderm, on each side of the head This ingrowth becomes separated from the surface to form a hollow ball of cells the auditory vesicle A constriction of the vesicle partially subdivides it into a dorsal utriculus and a ventral sacculus From the utricular part there arise three flat pockets whose central portions fuse and disappear, the remaining peripheral portions constitute the semicircular canals An enlargement of each canal forms the ampulla in which the crista differentiates The developing labyrinth influences the surrounding embryonic connective tissue to differentiate into cartilage which is later transformed into bone with the resultant formation of the bony labyrinth Between the membranous labyrinth and the cartilage there is connective tissue in early stages This connective tissue tends to disappear during development and to be replaced by the fluid perilymph This change occurs especially in vibration conducting channels leading from the oval window to auditory sensory areas The human labyrinth is one of the few structures that grow little if at all after birth

In fishes there is no middle ear, and the gill arches are used for gill respiration In higher forms, however the gill arches and associated structures are utilized during development in the formation of middle ear structures The gill cleft between first and second gill arches fails to communicate with the exterior in developing mammals, and enlarges to form the auditory tube and tympanic cavity The tensor muscle of the tympanic membrane is derived from the first embryonic gill arch, or mandibular arch, and the stapedius muscle

is derived from the second embryonic or hyoid gill arch *See* GILL (ANATOMY)

The quadrate cartilage of lower vertebrates is much reduced and persists as the incus of the adult mammal The proximal part of Meckel's cartilage the jaw of lower forms persists as the malleus and so the primitive jaw joint of reptiles and birds persists as the joint between incus and malleus in the adult mammal The upper part of the second gill arch skeleton persists and develops as the columella stapes of reptiles and birds

The external auditory meatus is derived from an ingrowth from the outer layer (ectoderm) within which a cavity develops The tissue between this cavity and the developing tympanic cavity differentiates into the tympanic membrane

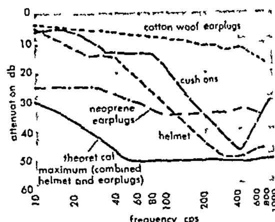
In the vertebrate series from fish to mammal there is a gradual shifting of the sound wave transmitting mechanism from medial to lateral side of the internal ear from cranial cavity to tympanic membrane and its associated ossicles In amphibians and reptiles both types of transmission are found *See* PHONORECEPTION SENSE ORGAN *see also* DEAFNESS [F K H]

Bibliography T H Bast and B J Anson *The Temporal Bone and the Ear* 1949 J E V Boas *Äusseres Ohr* in L Bolk E Göppert E Kallius and W Lubosch (eds) *Handbuch der vergleichenden Anatomie der Wirbeltiere* vol 2 part 2 1934 W M Copenhaver and D D Johnson *Bailey's Textbook of Histology* 14th ed 1958 H M de Burlet *Die innere Ohrsphare die mittlere Ohrsphare* in L Bolk E Goppert E Kallius and W Lubosch (eds) *Handbuch der vergleichenden Anatomie der Wirbeltiere* vol 2 part 2 1934 A A Maximow and W Bloom *A Textbook of Histology* 7th ed 1957 W Möllendorff (ed) *Handbuch der mikroskopischen Anatomie des Menschen* vol 3 part 1 1927 A S Romer *The Vertebrate Body* 1949

Ear protectors

Devices used to protect the human ear from noise that may be injurious to hearing There are situations in industry in the military services and in commercial aviation where persons must be exposed to sounds having levels in excess of those which may damage hearing Under these conditions damage to hearing can usually be avoided by having the people exposed to the noise wear one or more ear protective devices There are three general classes of ear protectors (1) earplugs that fit into the ear canal (2) earmuffs or cushions that fit over the external ear under which earplugs may be worn and (3) rigid helmets that fit over the entire head under which earplugs and earmuffs may be worn

The amount of protection afforded by an ear protector or combination of protectors is measured in terms of the number of decibels it attenuates or reduces the intensity of a sound that is transmitted through it In general the higher frequency components in a noise are attenuated more than are the lower frequency components



Attenuation of sound caused by various ear-protective devices (After C M Harris)

The attenuation afforded by various types of ear protective device is shown in the accompanying figure In order to achieve the reductions in sound pressure levels indicated it is necessary to have a good airtight seal between the ear protector and the surfaces of the ear or head against which it bears

Ear protectors reduce the level of speech or other wanted signals to the same degree that they reduce the noise Usually however, the reception and understandability of the speech or other signals is not adversely affected Signal reception is not adversely affected because the signal-to-noise ratio at the listener's eardrums is the same with or without ear protectors At very intense noise levels the attenuation of both the signal and the noise reduces overloading and distortion in the ear As a result speech intelligibility for example is better in intense noise when ear protective devices are worn than when they are not worn *See* NOISE CONTROL [K D K]

Bibliography C M Harris (ed) *Handbook of Noise Control* 1957

Earth

The third planet in the solar system lying between Venus and Mars and the only part of the known universe to present the intermingling conditions of air water and land making possible such a life zone as that at the face of this terrestrial globe Various scientific investigations are aimed at increasing knowledge of the inner surficial and outer parts of the earth and of their relationships with external parts of the universe Most of the topics outlined in this article lie within the general field of geophysical investigation (*see* GEOPHYSICS METEOROLOGY)

Shape gravitation, and density The earth is roughly elliptical in shape (Table 1) the equatorial bulge being approximately what would be expected for a rotating fluid Surface elevations are referred to a theoretical surface called the geoid, the equipotential surface which most nearly approximates mean sea level It can be visualized as the surface passing through the continents which

Table 1 Dimensions of the earth*

Equatorial radius	6 378 099 ± 116 m
Polar radius	6 356 631 m
Rad us of sphere of equal volume	6 371 200 m
Ellipt city	0 00336>9 ± 0 0000011
Volume	$1\ 083 \times 10^{t}$ cm³
Mass	$5\ 975 \times 10^{27}$ g
Average density	5 517 g cm³
Moment about polar axis, A	$8\ 0_3 \times 10^{44}$ g cm²
Moment about equatorial axis, C	$8\ 08 \times 10^{\ 4}$ g cm²
$(C - A)/A$	0 003273

* B Cutenberg *Internal Constitut on of the Earth* Dover 19₅1 H Jeffreys *The Earth* Cambridge 19₅2

the water table would assume if the rock were a porous permeable mass which exerted no forces on the water other than purely gravitational forces The geoid departs from the approximating spheroid probably by 100 m at most

The shape of the earth is commonly specified in terms of its gravitational field

$$g = g_0[1 + a \sin^2 \varphi + b \sin \ 2\varphi + c \cos^2 \varphi \cos 2(\lambda + \lambda\ )]$$

where g_0 a b c and λ are constants φ is latitude and λ is longitude measured westward from Greenwich (Table 2) The earth closely approximates a spheroid of revolution in which case c is zero Computations based on the shapes of satellite orbits suggest that additional terms may be needed to

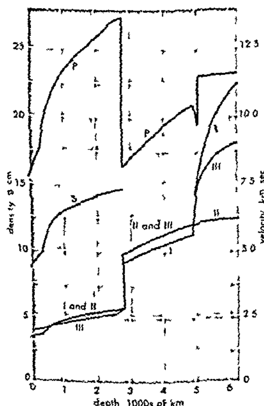


Fig 1 Three models of the dens ty variat ons w th depth (I II III) and the velocity of d latat onal (P) and shear (S) waves (After K E Bullen Introduct on to the Theory of Se smology 1947)

describe accurately the earth s shape Quantitative figures for the pear shapedness or other deviations of the earth s figure are not yet established See GEODESY

The average density of the earth is 5 517 g/cm³ Because the density of surface rocks is only 1 6-3 4 g/cm the interior of the earth must consist largely of rocks of greater density (Fig 1) This is indicated also by the earth s moment of inertia A uniform sphere would have a moment of 0 4 MR where M is its mass and R its equatorial radius The earth s moment is 0 3337 MR^2

Table 2 Values of constants in the gravity formula*

g_0	a	b	c	λ	Source
9 8 0490	0 005*88	0 0000059	0		International Formula of 1930
9 8 0596	0 00 2934	0 0000059	0		Uot la 1957
978 0*16	0 00₅2910	0 00000₅9	0 0000106	6°	Uotila 1957

* W A He skanen and F A Ven ng Me nesz *The Earth and its Grav ty Field* McGraw Hill 19 8

Water and land distribution Water covers 70 8% of the earth s surface (Table 3) The land is very unequally distributed on the globe Over 80% lies in the hemisphere centered at 38°N lat 0° l ng Furthermore the land is predominantly in the Northern Hemisphere (Fig 2) most of it ly ing in south pointing wedges around the Arctic Basin The proportion of land over sea surface is at a maximum at 66°N lat The land itself has a generally concentric arrangement with mountain ranges fringing the continents and old crystalline rocks outcropping in the centers The Himalayan region of Asia is the highest land Mount Everest (8840 m) being the highest peak

The oceans may be visualized as a series of basins separated by ridges The deepest parts of the oceans are not in their centers but in narrow troughs along their edges or adjoining submarine ridges and island arcs (Table 4) See OCEANS AND SEAS SUBMARINE TOPOGRAPHY

The ocean floors are usually covered by 1 3 km of sediments From them rise mountains of volcanic rocks in whole ranges lines of cones or isolated peaks These volcanic rocks appear to be different chemically from the typical rocks of the continents (Table 5) The oceanic volcanoes produce only basaltic lavas whereas both basalts and more acidic extrusives are found on the continents The acidic rocks are lighter in weight than the ba

Table 3 Area and elevat on of the land and oceans

Earth d v s ons land and water	Area % of earth	Area km² × 10⁶	Average eleva-tion meters	Source
Land	*9 *	148 89*	840	Kossinna (19 1)*
Ocean	0 8	361 059	−3800	Koss nna (19*1)
Cont nental shelves	5 4	27 5	−100	Kossinna (19*1)
Cont nental slopes	9 8	50	−2*00	Howell (1950)
Ocean bas ns	47 8	*43 6	−4860	
Seas	7 8	39 9 8	−1*10	Kossinna (19*1)
Whole earth	100 0	510 1	−*440	Kossinna (1921)

* E K ssinna D e T efen des Weltmeeres Veröffentl Inst Meereskunde Un v Berlin Neue Folge A Geogr -na 9 1 70 19*1

Table 4 Greatest depths in the oceans*

Name	Depth in water meters	Ocean basin	Lying near
Challenger Deep	10 863	Pacific	Mariana Islands
Mindanao Deep	10 497	Pacific	Philippines
Ramapo Depth	10 374	Pacific	Honshu Japan
Tonga Kermadec Trench	10 035	Pacific	Tonga Kermadec Islands
Planet Depth	9 140	Pacific	New Britain
Milwaukee Depth	8 750	Northwest Atlantic	Puerto Rico
Bonin Trench	8 660	Pacific	Bonin Islands
Byrd Deep	8 5[illegible]0	Pacific	Southeast of New Zealand
Tuscarora Depth	8 500	Pacific	Kurile Islands
South Sandwich Trench	8 264	Southwest Atlantic	South Sandwich Islands
Aleutian Trench	7 680	Pacific	Aleutian Islands
Atacama Trench	7 635	Pacific	Northern Chile
Ryukyu Trench	7 480	Pacific	Ryukyu Islands
Sunda Trench	7 455	East Indian	Java

* From G P Kuiper (ed) *The Earth as a Planet* Univ of Chicago Press 1954 H U Sverdrup M W Johnson and R H Fleming *The Oceans* Prentice-Hall 1942

saltic and the height of the continents is thus explained The surface of the earth can be looked upon as being divided into two predominant levels (Fig 3) the ocean basins at a depth of 4–6 km and the continental plateaus from −200 m to +1 km Above these stand the mountain ridges below them dip the ocean deeps They are separated by the abrupt (2 3 5°) drop of the continental slopes

Enveloping atmosphere The earth is surrounded by an envelope of gas the atmosphere This is arbitrarily divided into layers on the basis of temperature (Fig 4) The lowermost layer the troposphere contains about three-fourths of the mass of the atmosphere and is the only layer capable of supporting life as known on earth It varies in thickness from about 8 to 17 km bulging at the Equator The tops of even the highest mountains are all within the troposphere as are most of the cloud and all of the weather experienced on the ground *See* ATMOSPHERE

At the critical level of 450 550 km the horizontal mean free path of the gas particles (distance between collisions) becomes equal to the elevation and the gas is so tenuous that the term tempera

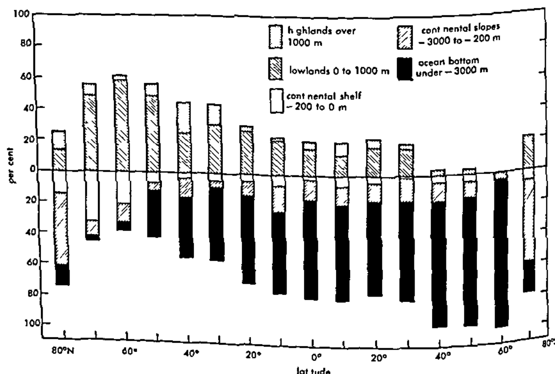


Fig 2 Variation of elevation with latitude along selected parallels (Howell 1959)

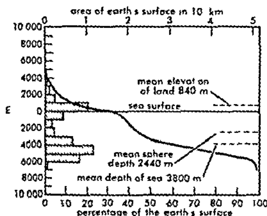


Fig 3 Hypsographic curve of land surface elevation (After E Kossinna Die Tiefen des Weltmeeres Veroffentl Inst Meereskunde Univ Berlin Neue Folge A Geogr-naturwiss 9 1–70 1921)

ture ceases to have its conventional meaning Above this point light atoms such as hydrogen and possibly helium if given sufficient thermal energy can escape from the earth altogether (escape velocity ≈ 11 2 km/sec)

The atmosphere becomes increasingly ionized with elevation Because of the high concentration of electrically charged particles the upper atmosphere (ionosphere) is a good conductor of electricity The electrical charges move in loops in the earth s magnetic field and this motion tends to restrain particles in the earth s neighborhood far beyond the critical level *See* AERONOMY IONOSPHERE

Table 5 Average composition of rocks believed typical of the major layers of the earth wt%

Element	Continental crust*	Plateau basalt†	Stone meteorites‡	Iron meteorites‡
O	46 59	44 67	36 15 ± 0 89	
Si	27 72	22 81	18 12 ± 0 22	
Al	8 13	7 40	1 53 ± 0 13	
Fe	5 01	10 11	24 18 ± 1 08	90 78 ± 0 26
Ca	3 63	6 70	1 74 ± 0 18	
Na	2 85	1 97	0 69 ± 0 05	
K	2 60	0 57	0 18 ± 0 02	
Mg	2 09	4 04	13 93 ± 0 29	
Ti	0 63	1 31	0 08 ± 0 03	
H	0 13	0 20	0 06 ± 0 02	
P	0 13	0 14	0 14 ± 0 01	
Mn	0 10	0 13	0 26 ± 0 06	
S	0 052		1 79 ± 0 08	
C	0 032			
Cl	0 048			
Cr	0 047		0 30 ± 0 05	
Ni	0 020		1 53 ± 0 16	8 59 ± 0 24
Co	0 001		0 10 ± 0 02	0 63 ± 0 02
Other	0 200			

* B Gutenberg (ed) *Internal Constitution of the Earth* 1951

† R A Daly *Igneous Rocks and the Depths of the Earth* 1933

‡ H Brown and C Patterson The composition of meteoric matter II *J Geol* 56 85–111 1948

Temperatures air and earth At the surface of the earth the average annual temperature varies from around 90°F (32°C) to lower than −25°F (−32°C) Among the highest and lowest recorded temperatures (in shade) are −125°F (Vostok Antarctica) and +136 4° (Libyan Desert) The gradient with depth at the surface generally lies in the range 0 01–0 04°C/m Figure 5 is one estimate of the possible temperature variation to 800 km The average heat loss at the surface by conduction

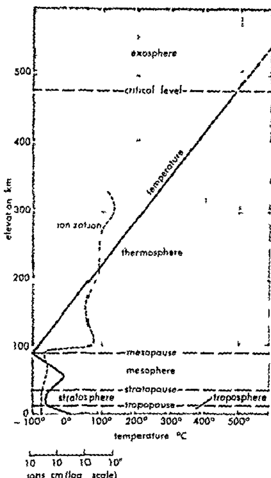


Fig 4 Structure of the atmosphere

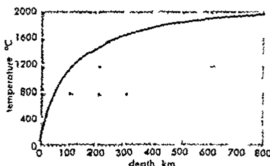


Fig 5 Temperature as a function of depth (Howell 1959)

from below is about $1.2 \times 10^{-6} \pm 50\%$ g cal/(cm²)(sec) This is equivalent to 6×10^{12} cal/sec for the whole earth *See* EARTH (HEAT FLOW) The additional heat loss through volcanoes and hot springs is at least two orders of magnitude smaller The estimated concentrations of radioactive elements in the earth s crust are sufficient to provide all this heat by their disintegration Therefore it is believed that their concentration decreases rapidly with depth at least under the continents

Internal crustal patterns Most information about the earth s interior is provided by seismology From a study of the arrival times of seismic pulses it is known that the earth contains three main layers the crust the mantle and the core (Fig 6) The composition of the crust beneath the continents is believed to vary with depth Much of the surface is covered by a thin veneer of sedimentary rocks 0–3 km thick In a few places great th kne e n 20

ature and pressure The rocks of the continents are largely silicates aluminum rich at the surface (sial) grading to iron and magnesium rich at depth (sima) The sial is missing beneath the oceans It is uncertain how sharp a boundary exists between the sial and sima There is probably an intermediate zone in which the velocities of earthquake waves are at a minimum Seismic pulses are not returned directly to the surface from this zone hence little is known of its nature

At the lower boundary of the crust a sudden increase in seismic wave velocity occurs This is known as the Mohorovičic discontinuity after its discoverer *See* MOHO (MOHOROVIČIC DISCONTINUITY) It may represent a change in composition or only a phase change from one suite of mineral species to another of identical chemical composition No direct evidence of the composition of the mantle or core exists The increase in velocity at the Mohorovicic discontinuity is consistent with what one would expect if the composition changed from a gabbro to a more ultrabasic rock such as dunite or eclogite On the assumption that meteorites are a sample of a fragmented former planet like the earth it would be expected that the bulk of the earth is composed of ultrabasic silicates and metallic iron (Table 5) According to this analogy the outer mantle would be composed of silicates The transition zone starting with the second-order seismic discontinuity at about 412 km (Fig 1) could represent the introduction of some nickel iron into the system and the core would be all nickel iron The 412 km discontinuity could also be due to a phase change

In the crust and the outermost part of the mantle most rocks are brittle and break before they will flow plastically Studies of the variation of gravity over the earth's surface in relation to its history of erosion and sedimentation indicate that the surface rises and falls to maintain a sort of hydrostatic balance called isostasy (*see* TERRESTRIAL GRAVITATION) This implies *that the rocks*

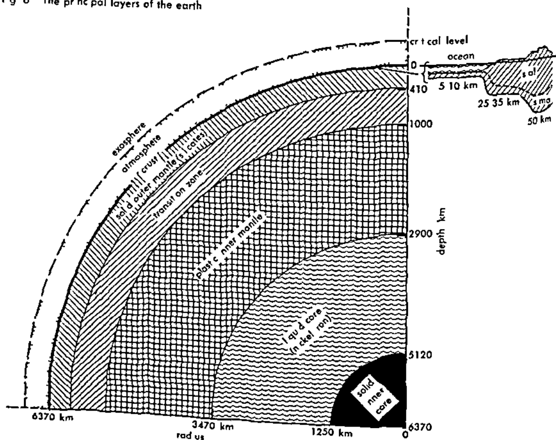


Fig 6 The principal layers of the earth

beneath some depth (96 km) called the depth of compensation are plastic and of low strength How ever there is a finite breaking strength at least to a depth of 700 km because earthquakes apparently of shear origin occur to that depth

The surface of the earth is continually undergoing deformation All coast lines show evidence of repeated rise and fall of the land relative to the sea Some of this deformation but not all is due to changes in sea volume Sea level falls of 75–100 m have occurred when water was locked in continental glaciers (*see* GLACIER) If all the present glaciers were melted the sea level would rise roughly 60 m The volume of surface water is presumably being increased slowly by the new water in volcanic gases The amount is so small that sea level has not been increased noticeably in the last 5×10^9 years as proved by the distribution of sediments One might expect that in this length of time erosion would have worn away all the continents Some process acts to preserve the continents causing them to rise repeatedly over large areas and to be thrown into mountains in long narrow belts The forces causing these changes must come from the earth s interior One theory is that the interior of the earth is shrinking in size causing the crust to wrinkle as it adjusts Another is that subcrustal currents sweep parts of the crust together producing thick welts of mountain ranges *See* OROGENY

Both dilatational and shear waves are transmitted throughout the crust and mantle In the outer part of the core however no shear waves are observed suggesting that it is fluid The change from the solid to liquid state is sufficient to account for the decrease in seismic velocity here (Fig 1) but the core s great density suggests that there is a change in composition also It has been suggested that the core is a plasma a material so compressed that the orbital electrons have been collapsed around the nuclei of the atoms If this is the case the composition may be largely hydrogen and helium as in stars However it is doubtful if sufficient pressure exists to create a plasma of this sort within the earth Within the core is a central body of higher seismic velocity than that of the outer part This inner core may be solid *See* EARTH INTERIOR SEISMOLOGY

Major planetary relationships The earth moves about the sun in an elliptical orbit (eccentricity = 0 01674 average radius = $1\,495 \times 10^8$ km) in a little over a year The tropical year on which our calendar (Gregorian) is based is the period between vernal equinoxes which precess in position with respect to the stars with a period of 25 800 years The *rate of precession* 50 2 *sec/year* is not constant but has a fluctuation called nutation of about 9 23 sec Both precession and nutation are due to the gravitational attraction of the sun and moon on the earth s equatorial bulge

The earth rotates 365 2422 times on its axis over a period of a year This period is not constant however but is increasing at a rate of a little more than 1 msec/100 years The cause of this change is a transfer of energy from the rotation of the earth to the revolution of the moon about the earth through the action of tidal friction There are also irregular fluctuations of as much as 5 msec in the length of a day The best explanation of these is that the earth s core is in turbulent fluid motion This motion is coupled to the mantle and crust of the earth in such a way that changes in the surface motion compensate for changes in the core currents to keep the total moment of inertia constant

The axis of the earth is tilted 23°26 59′ to the plane of its revolution The axis of rotation does not coincide with the axis of figure but circles about it in a counterclockwise direction with a maximum separation of about 0 4 sec The period of this Chandler (Eulerian) motion is about 14 months In addition it is believed that the position of the axis of rotation of the earth (or of the crust with respect to the interior) may have shifted greatly in geologic time (*see* POLAR WANDERING) Evidence for such changes comes from studies of the direction of remnant magnetization of rocks *See* GEOGRAPHY MATHEMATICAL ROCK MAGNETISM

Satellites The earth has one natural satellite and a variable number of artificial ones (Table 6) The natural satellite is the moon which moves about the earth in an elliptical orbit at a mean distance of 383 403 km with an eccentricity of 0 05490 and a period of 27 days 7 hr 43 min 11 5 sec Its orbit has an average inclination to the orbit of the earth of 5°8 33 Actually the earth and moon move about a common center of gravity whose motion about the sun is more regular than that of the earth Because of the lesser mass of the moon

Table 6 Earliest long lived artificial satellites

Name	Launched	Perigee km	Apogee km	Eccentricity	Orbital inclination degrees	Period min	Weight kg
58 β 1 Carrier	March 17 1958	7030	3860	0 20744	34 25	138	22
58 β 2 *Vanguard* I	March 17 1958	7030	3970	0 1895	34 25	134	1 4
59 α 1 *Vanguard* II	February 17 1959	6940	3360	0 16554	32 86	126	9 8
59 α 2 Carrier	February 17 1959	6940	3380	0 18378	32 88	130	—

(^{1}/51 8 of that of the earth) this center is within the earth at an average radius of 4645 km

Theories on age and evolution Many lines of evidence suggest an age for the earth of $4–6 \times 10^9$ years The presence of any radioactive elements in the earth at all indicates that the material of which it is formed was created under entirely different conditions from any found in the earth to day The only places known today where radioactive elements are likely to form are the centers of exceptionally hot dense stars Assuming all the heavy nuclides were formed in nearly equal quantities as seems likely their relative abundance suggests an age for these elements of about 6×10^9 years *The* relative amounts of radioactive elements and their end products in the earth suggests an age *of* $5\ 5.5 \times 10^9$ years for the earth s crust The oldest rocks found anywhere on the earth appear to have been formed $2\ 3.5 \times 10$ years ago Meteorites have ages (*from studies* of their radioactivity) of about 4.5×10^9 years *See* DATING METHODS GEOCHRONOMETRY

The rate of recession of distant galaxies has been estimated to be such that about 5×10^9 years ago all were grouped tightly in a small volume of the universe The origin *of the earth* may thus have been a minor event in the origin of the galaxy or *of all* known galaxies

At one time it was thought that the earth was a *fragment* of a larger star which had somehow been disrupted A more widely held hypothesis is that the sun and all bodies *of the* solar system condensed from a dispersed gas The smaller planets *lacked strong* enough gravitational fields to retain their hydrogen and helium and thus they differ in composition from the larger planets and the sun *According to one* variation of *this theory* the earth formed originally by coagulation of cold particles in a turbulent gas cloud At first the earth was uniform in composition but as *it grew in* size the temperature of the interior *increased* through compression and radioactive decay *and iron* separated from the silicate phase to form the core The bulk of the elements particularly *the light* and the radioactive elements was concentrated *upward* This process is presumed to be still going on Radiogenic heat and gravitative *forces are* the two major sources called on to explain the deformations to which the crust is continually *subject* The thermal conductivity of *the earth is* so low that it is possible that the earth s interior *may still* be warming It is believed by some on geologic grounds that the rate of change of *the earth's* surface features has been accelerating throughout *its* history *In* any case the earth is still undergoing change and can be expected to evolve for some time to come *See* COSMOCHEMISTRY COSMOLOGY

[B F HO]

Bibliography K E Bullen *Introduction to the Theory of Se* ... Geophys Suppl 6 50–59 1950 R A Daly *Igneous Rocks and the Depths of the Earth* 1933 B Gutenberg *Internal Constitution of the Earth* 2d ed 1957, W A Heiskenen *Size and Shape of the Earth,* Inst Geodesy Photogram Cartog Publ 7, 1957, B F Howell, Jr, *Introduction to Geophysics* 1959, H Jeffreys *The Earth its Origin History and Physical Construction* 3d ed 1952

Earth (age of)

The age of the earth may be defined as the time that this planet has existed with approximately its present mass and density Today the best estimate for this period is 4.5×10^9 years

Once the general physical features of the earth and its solar environment became known the age of the earth was determined by means of long term and steady processes whose rates and extent of progress could be measured Such processes as the cooling of the earth the accumulating of sediments and the salting of the oceans were used Today the process of radioactivity is utilized where the time required to form a measured amount of decay product associated with a radioactive element is calculated from the known rate of disintegration of the parent element

Minimum age It is difficult to apply this method to the whole earth since any rock chosen as a sample for study is a secondary system younger than the earth For this reason the age of the oldest rocks that can be sampled will give only a minimum age for the earth In every continent there are limited regions which have somehow escaped pervasive destruction by weathering and igneous activity for very long periods of time These oldest regions called shield areas because of their stability, consist of rocks little different from any formed during subsequent times The ages of minerals in some of these oldest available rocks have been determined by various radioactive decay systems such as $K^{40} \rightarrow A^{40}$, $Rb^{87} \rightarrow Sr^{87}$ or $U^{238} \rightarrow Pb^{206}$ as about 3.0×10^9 years *See* ROCK (AGE DETERMINATION)

Maximum age A maximum age for the earth is given by theories for the cosmogenic origin of the elements According to present theory the nuclides U^{235} and U^{238} were formed in the ratio 1.64 1 within an ancient star just before it exploded and scattered debris in space Our sun and earth were subsequently formed of material which included this debris U^{235} and U^{238} are both radioactive but U^{235} decays much faster than U^{238} so that the ratio of U^{235} to U^{238} is constantly decreasing The U^{235}/U^{238} ratio is 0.007 in the earth today If one calculates how far back in time one must go in order to increase this ratio to the maximum prescribed by cosmogenic theory one finds that it takes 6.6×10^9 years Since the earth is younger than any element formed in this manner the maximum age of the earth from this point of view is 6.6×10^9 years Corrections must be made for the production of uranium in stars exploding during a finite interval or at a constant rate but a maximum age is not seriously increased by the choice of a reasonable cosmological model

Ore lead age. The ratio of radioactive parent and accumulated daughter of a decay scheme cannot be measured within the whole earth, but it is possible to make a direct measurement of the age of the earth by comparing the relative progress of the two radioactive decay schemes $U^{235} \rightarrow Pb^{207}$ and $U^{238} \rightarrow Pb^{206}$ This group of four nuclides has an important and singular property The ratio of radiogenic Pb^{207} to Pb^{206} formed during a period of uranium decay is time-dependent, and the age of any closed system containing uranium is easily determined by measuring only the ratio of radiogenic Pb^{207} and Pb^{206} contained in it, since the different decay rates of the uranium nuclides and their relative abundances are known *See* LEAD ISOTOPES GEOCHEMISTRY OF

Of all the lead in the earth only part of the Pb^{206} and Pb^{207} has been formed by uranium decay because there was some primordial lead Pb^{206} and Pb^{207} in the earth when it was formed A very useful characteristic of primordial lead is that it contains Pb^{204}, whose abundance in the earth does not change for it does not have a long lived radioactive parent As a consequence, changes in the abundances of Pb^{206} and Pb^{207} may be determined simply by comparing their abundances against Pb^{204} In primordial lead the Pb^{206}/Pb^{204} and Pb^{207}/Pb^{204} ratios have fixed values When radiogenic Pb^{206} and Pb^{207} are added, these ratios increase to higher values

Nearly all lead in the earth is a mixture of primordial lead and radiogenic lead At any given time there is only a small quantity of radiogenic lead which exists separately in uranium minerals These minerals are continuously being formed and destroyed so that their radiogenic leads are continuously being mixed into the common pool of earth lead The radiogenic component of common earth lead is therefore the sum of many small increments added during the earth's lifetime

If a uranium mineral contains primordial lead then the radiogenic Pb^{206} which has accumulated may be found by subtracting the primordial Pb^{206}/Pb^{204} ratio from the total Pb^{206}/Pb^{204} ratio in a sample of lead from the mineral The radiogenic Pb^{207} may be similarly found The ratio of these two radiogenic uranium leads is the time dependent Pb^{207}/Pb^{206} ratio useful for age calculations In a practical case if two minerals have the same age and both contain primordial lead but only one contains uranium then the age can be readily calculated from the ratio of the differences between the uranium leads in the two lead samples Similarly, if two minerals have the same age and both contain primordial lead and both contain uranium but in different proportions then the age can be calculated as before from the ratio of the differences between the two leads without knowledge of the composition of primordial lead The first truly significant measurement of the age of the earth, the ore lead method was based on this principle Common earth leads are obtained from lead ore minerals which exist separately in small amounts just as uranium minerals do The earth is such a vast object that the lead from one continent may never mix well with that from another and if there were more uranium or less primordial lead in one continent than another, their common leads might contain different proportions of radiogenic lead The age of the earth might be calculated from the ratio of the differences between common ore leads in the two continents The actual differences among common earth leads are so small that many different lead samples and statistical approaches must be used in practical calculations The age obtained in this manner was about 3×10^9 years There are many theoretical difficulties involved in this method but by using a more reliable age of the earth (as given below) this same type of calculation can be used to study the mixing of rocks within the earth

Meteorite lead age The most acceptable way to measure the earth's age is by the meteorite lead method This is simply an extension of the ore lead method to include meteorites When leads are isolated from various iron and stone meteorites and their different isotopic compositions are compared a simplified pattern emerges the Pb^{206}/Pb^{204} and Pb^{207}/Pb^{204} differences among the leads cover a very large range but the Pb^{206}/Pb^{207} ratio obtained from the Pb^{206}/Pb^{204} and Pb^{207}/Pb^{204} differences between any two leads is constant The age calculated from this constant is 4.5×10^9 years and for meteorites has been confirmed in a general way by other less accurate radioactive decay methods that is meteorites may be considered separate little planets all formed at the same time all containing the same kind of primordial lead and all containing varying proportions of radiogenic lead Because some meteorites contain essentially no uranium enough information is available not only to calculate an accurate age for meteorites but to describe and define completely all the possible meteoritic leads which can exist

These meteoritic leads constitute only a very small fraction of the total of every possible kind of lead so that if the lead from an object of unknown age fits the description of a meteoritic lead it is probably safe to conclude that the object has the age calculated for meteorites This is the case for the earth Samples of common lead are obtained from the oceans which contain mixtures of leads derived from all lands and these leads fit the description of meteoritic lead One may therefore conclude that the earth has the same age as meteorites namely 4.5×10^9 years *See* GEOLOGICAL TIME SCALE, METEORITE [C C P]

Bibliography E M Burbidge G R Burbidge W A Fowler and F Hoyle Synthesis of the elements in stars *Revs Modern Phys*, 29 547 1957, C Patterson Age of meteorites and the earth *Geochim et Cosmochim Acta*, 10 230 1956

Earth (heat flow)

Terrestrial heat flow or the amount of thermal energy escaping from the earth per unit area and unit time is of fundamental importance to geolo- geophysics The heat leading to the fusion

amorphism of rocks and the forces of mountain building are thought to originate in deep seated thermal processes A most useful quantity both for the general theory of these processes and for the es timation of temperatures beyond accessible depths is the terrestrial heat flow

Heat flow can be measured at the surface of the earth but it is expected to decrease with depth This is because the thermal flow at shallow depths originates in part in radioactive decay at moderate depths and in the cooling of shallow rocks The heat flow at great depths is unaffected by these processes It is necessary to have information about the distribution of radioactive heat sources and about changes in temperature to infer the heat flow at great depths

A major goal of geothermal investigations is the deduction of temperature throughout the earth This can be found from the heat flow if the ther mal conductivity is known The transfer of heat in solids at the high temperatures and pressures thought to prevail in the earth may take place by mechanisms which are unimportant under the con ditions of ordinary experience and care must be exercised in the extrapolation of laboratory experi ments to these extreme physical conditions

Heat flow at the surface This is determined by multiplying the rate of increase of temperature with depth (the geothermal gradient) by the local thermal conductivity The latter quantity is usu ally measured in the laboratory on samples taken from the site of the temperature observations The geothermal gradient can be determined on land wherever undergound temperatures can be meas ured It is usually necessary to have data extending to a depth of at least 1000 ft in order to avoid disturbances from diurnal and annual temperature fluctuations climatic change and circulation of ground water Suitable measurements have been made in boreholes mines and tunnels Heat flow through the ocean floor is measured by dropping a metal probe equipped with temperature sensing de vices into the sea bottom Samples for measurement of thermal conductivity are collected by taking cores nearby *See* BORING AND DRILLING MINERAL

Heat flow has been measured in about 25 places on land and in about 100 at sea Results on land range from 0 7 to 2 5 × 10^{6} cal/(cm^2) (sec) and those at sea from 0 2 to 8 × 10^{6} cal/(cm) (sec) The number of determinations is still too small to permit analysis of the dependence of heat flow on locality The oceanic values have a wider range than the continental this is probably due in part to encountering regions of recent volcanism in some of the oceanic measurements

Average heat flow is about 1 2 × 10^{-6} cal/(cm^2) (sec) which corresponds to a loss of about 2 × 10^{0} cal/yr for the whole earth The uncer tainty in this figure is about 50% The loss of heat by conduction apparently greatly exceeds that lost by other processes such as volcanism but it is very small compared with the amount of solar radiation absorbed and reradiated Its effect on the surface temperature is negligible

Radioactive heat production The important heat producing elements in the earth are uranium thorium and potassium The heat produced by the uranium series is roughly equal to that produ ed by the thorium series in most natural material Heat production in igneous rocks ranges from about 600 × 10^{15} cal/(cm^3) (sec) in granites to 0 2 × 10^{15} cal/(cm^3) (sec) in dunites It is asso ciated with the tenor of feldspar and silica be ng highest in the light colored low density rocks and low in the dense dark colored rocks Heat produc tion in stony meteorites is about 3 5 × 10^{-15} cal/(cm^3) (sec) In the latter case about 60% of the heat results from the decay of K^{40} in rocks the heat from potassium is roughly one fourth to one third of the total

Measurements of heat flow show immediately that rocks having the radioactivity of those making up most of the surface of the earth must be con fined to a thin shell Otherwise their heat produc tion would cause a higher heat flow than is ob served This implies strong upward concentration of the radioactive elements

Further considerations about the distribution of radioactivity with depth stem from the apparent equality of continental and oceanic heat flow The measurements of heat production imply that half or more of the continental heat flow originates in crustal rocks Since the crust in the ocean basins is comparatively thin and apparently devoid of gra nitic rocks its heat production must be corre spondingly lower The simplest explanation of the constancy of heat flow is that the total amount of radioactivity beneath unit area of the earth s sur face is roughly the same everywhere On this inter pretation the difference between the radioactive distributions beneath continents and oceans is that there is more radioactivity in the mantle in oce anic regions The mantle beneath the continents has presumably been impoverished in radioactive elements during the formation of the continental crust

The higher concentration of radioactivity at depths beneath the oceans implies that the temper atures there are higher than beneath the conti nents This difference need not be large if heat pro duction is concentrated near the surface

It is commonly supposed that the earth has the same bulk composition as meteorites If its average content of radioactive elements is the same as the stony meteorites its total heat production is about 2 3 × 10^{0} cal/yr compared with a total heat l s of about 2 × 10^{0} cal/yr These figures do not d f fer significantly and they imply that substantially all of the heat being produced is escaping on this model This in turn implies that most of the heat sources must be close to the surface since heat produced at depths greater than about 1000 km has not had time in the age of the earth to reach the surface

It cannot be concluded however that all of the heat escaping today originates in present radioac tive decay Modern thinking is overwhelmingly in favor of the idea that the earth accumulated at a

relatively low temperature and heated during its early history Hence little or no initial heat could be escaping at present But the heat production in the earth was higher in the past because of radioactive decay and heat that was produced by nuclides which decayed long ago may make an appreciable contribution to the present heat flow Estimates suggest that 50% or less of the present heat flow at the surface could have originated in this way but this figure is large enough to cast doubt on the earth model based on meteorites

Heat transfer in the earth. All available evidence suggests that temperatures of a few thousand degrees must be expected in the earth Few materials remain solid at these temperatures unless they are subjected to high pressure and mechanisms of heat transfer which are unimportant under common laboratory conditions may be very important in the earth Effects of pressure must also be considered

Heat conduction This takes place in solids in two basically different ways Conduction by phonons (lattice vibrations) is the only important mechanism of heat transfer in dielectric solids at room temperature in metals conduction by mobile charged particles dominates Both types of conduction take place in the earth

Phonon conductivity in dielectric solids at high temperatures is usually observed to decrease with temperature according to the relation $K = (AT + B)^{-1}$ where K is conductivity T is temperature and A and B are adjustable constants Its variation with pressure is virtually unstudied experimentally but a few theoretical estimates have been made The results indicate that the effect of pressure may nullify or even reverse the decrease in conductivity resulting from increasing temperature in the earth but this conclusion must be considered tentative until experimental data become available If it is correct the conductivity due to phonons would be in the range 0 005–0 01 cal/(cm) (sec) (°C)

Additional heat transfer takes place in electrically conducting substances In a metal the relation between electrical and thermal conductivity is given approximately by the Wiedemann Franz law which states that the ratio of the thermal conductivity to the electrical conductivity multiplied by the temperature is constant Application of this relation to the earth s core leads to a thermal conductivity of about 0 2 cal/(cm) (sec) (°C) but this numerical value depends on the electrical conductivity assumed and is subject to attendant uncertainties

In the semiconducting mantle the Wiedemann Franz formula must be modified to take account of the transport of excitation energy as well as kinetic energy (see WIEDEMANN FRANZ LAW) The resulting thermal conductivity however is unlikely to exceed 0 001 cal/(cm) (sec) (°C) in the lower mantle if the electrical conductivity σ equals 1 ohm^{-1}cm^{-1} This is about one order of magnitude less than the value expected for the phonon conductivity but if $\sigma = 10$ ohm^{-1}cm^{-1} which is probably within the uncertainty of the electrical data in the earth account must be taken of this effect

The transfer of heat by excitons may be more important Excitons are electron hole pairs bound together by their Coulomb attraction and as such they cannot contribute to electrical conduction They can however carry thermal energy Evidence for their existence in oxides and silicates is scanty and the importance of this process of heat transfer in the earth is highly uncertain

Radiative transfer Such transfer may become very important in non opaque solids at high temperatures The contribution of radiation to the thermal conductivity is given with ample accuracy by the expression $K = 16n^2sT^3/\epsilon$ where n is refractive index s is the Stefan Boltzmann constant, and ϵ is the sum of the absorption and scattering coefficients averaged over all wave lengths The other symbols have their previous meanings

The mean refractive index in the mantle may be estimated from the Gladstone-Dale law which is reasonably well satisfied by minerals it predicts a linear relation between index and density and the latter may be regarded as known in the earth No such direct procedure is available for estimating ϵ It appears however that absorption rather than scattering is important in limiting the mean free path of photons at most frequencies

Four processes leading to absorption of light in semiconducting solids may be distinguished Photon phonon interactions produce strong absorption in the infrared but this is unimportant because there is little radiant energy at these long wave lengths Optical excitation of electrons across the fundamental energy gap of the crystal produces strong absorption in the ultraviolet (visible in some cases) and such absorption may be important in the earth

In dielectric solids there is usually an interval of relatively high transparency in the visible and near infrared In many cases the absorption is sensibly zero but there are weak absorption peaks in crystals containing transition elements In the earth the most important transition element is iron It causes an absorption peak at a wave length of about 1 μ in addition to weaker absorption throughout the visible and near infrared

In electrically conducing materials absorption may take place at all wave lengths The relation between absorption coefficient α (in cm^{-1}) and conductivity (in ohm^{-1}cm^{-1}) is $\alpha = 120\pi\sigma/n$ All terms in this expression must be evaluated at the same wave length or frequency

The absorption between peaks is of paramount importance to radiative transfer Each spectral interval acts like an electrical resistor in a bank of resistances in parallel Only one perfectly transparent "window is required to give infinite thermal conductivity just as a perfect shunt leads to zero resistance in such a bank of resistors This analogy indicates the relative unimportance of the exact value of the absorption in the comparatively opaque intervals For this reason, the limitation of the mean free path by processes which are relatively independent of frequency are very important.

Room temperature measurements on a few ferromagnesian silicates have shown that the lowest absorption coefficients are a few cm^{-1}. These results are not directly applicable to the outer mantle because the absorption is likely to increase with temperature, and there are indications that it increases with pressure as well. The magnitudes of these effects are unknown, and until they have been investigated it seems unwise to consider an absorption coefficient less than 10 cm^{-1} in the mantle.

The electrical conductivity in the outer mantle is too low to lead to appreciable absorption, but at a depth of about 600 km the conductivity rises sharply to a value of about 1 $ohm^{-1}cm^{-1}$. High absorption may be associated with this feature of the earth because of the general absorption connected with electrical conductivity and the strong absorption at wave lengths shorter than the ultraviolet absorption edge. Evaluation of these effects depends on interpretation of the rise in electrical conductivity.

The electrical conductivity of a semiconductor may increase for two important reasons other than because of an increase in temperature. The drift mobilities (mean velocities in unit electric field) of the electrons and holes may increase, or the activation energy of conduction (which is proportional to the width of the fundamental energy gap) may decrease. If the rise in electrical conductivity is the result of an increase in drift mobility, the conductivity depends on frequency and decreases at high frequency. The conductivity in the earth is determined at essentially zero frequency, and if the frequency effect is large, the transparency may also be high. If, however, the energy gap decreases markedly at depth, the transparency is not as great. A smaller mobility implies a smaller frequency effect, and much of the visible part of the spectrum may be cut off at the ultraviolet absorption edge. The energy gap can hardly be expected to remain constant in any event, and experiment indicates that it decreases with pressure and temperature.

An alternative explanation of the rise in conductivity is that electricity is conducted by ions rather than electrons in the lower mantle. In this case the conductivity would have almost no effect on the transparency, since ions are too massive to interact with visible light.

Convective transport of heat. This circulation has been considered to take place in both the mantle and the core. The fluid outer core is thought to be in convective equilibrium, that is, the thermal gradient is thought to be that required to start convective motion. The amount of heat actually transferred by fluid motion may not be large, however, because of magnetically induced viscosity.

Convective transport of heat has also been suggested in the mantle despite seismic evidence for its rigidity. It is supposed that although the mantle behaves as a solid for short period stresses, it may be unable to support shears persisting for times of the order of 10^4 years.

If convection in the mantle leads to motions of a centimeter or so per year, and if the motion involves masses of material with dimensions of thousands of kilometers, then this process of heat transfer is dominant and the mantle is in convective equilibrium. But it is not clear that such motion could ever take place, since several arguments suggest that the mantle has a nonzero strength even for long continued stresses. In this case calculations indicate that the motion would be concentrated in filaments with a comparatively small amount of material involved. Convective transport of heat would be greatly reduced as a consequence, and it may not take place at all in the mantle. See EARTH, EARTH INTERIOR. [S.P.C.]

Bibliography: L. H. Ahrens, K. Rankama, and S. K. Runcorn (eds.), *Physics and Chemistry of the Earth*, vol. 1, 1956; G. P. Kuiper (ed.), *The Earth as a Planet*, 1954.

Earth (orbital motion)

With respect to the Sun, the Earth rotates on its axis, as demonstrated by a freely swinging pendulum, and revolves about the Sun, as demonstrated by the annual parallactic displacement of nearby stars against the background of distant stars. See FOUCAULT PENDULUM; PARALLAX (ASTRONOMY). Because the Earth, including its oceans and atmosphere, is not a symmetric rigid body and because it is not the only body revolving about the Sun, the motions vary with time. See EARTH; PERTURBATION (ASTRONOMY); TIME.

ROTATION ABOUT AXIS

Astronomical observations show that planet Earth undergoes three types of variations in speed of rotation: secular, irregular, and periodic. The secular increase in the length of the day, chiefly as a result of tidal friction due to the Moon, is about $0^s.0015$ per century. Irregular changes in speed are random and persist for several years. The total range in the length of the day during the past 200 years is nearly $0^s.01$. Periodic variations have periods of 1 year, 0.5 year, 27.55 days, and 13.66 days. These changes in the length of a day are on the order of $0^s.0005$.

Causes of variations. The speed of rotation of Earth is measured with respect to stations fixed on land. However, the total angular momentum of planet Earth is the sum of the momenta of land, water, and air. An interchange of momentum between these three will change the length of a day.

The crust of the Earth is elastic. Tides generated by Moon and Sun change the shape, changing the moment of inertia. The Earth may have a fluid core with turbulent motion. Changes in coupling between such a core and the crust would change the speed of rotation. Ocean and bodily tides decrease the angular momentum of the Earth through friction. The Earth may shrink or expand as a whole, or land masses may rise or fall with a change in moment of inertia.

Measurement of variation. Mean solar time is defined so as to be a strict measure of the angular rotation of the Earth. Hence variations in speed of rotation are obtained by comparing mean solar

time with some other form of time which is independent of the rotation of the Earth The following have been used for this purpose the orbital motions of the planets and satellites the oscillatory motions of the pendulum and quartz crystal and the vibrations of atoms as in atomic clocks

The practical construction in 1955 of an atomic clock of high precision opened up a new era in the study of the rotation of the Earth It has enabled periodic variations to be determined with high precision has permitted details of the irregular variation to be shown and will enable the secular variation to be determined with high precision in the future

In the past the variation in rotation has been studied principally by using the Moon Its angular motion with respect to the stars is the fastest of any celestial planet or natural satellite

Discordances between the observed and computed positions of the Moon could be due to imperfections in the lunar theory or to variations in the length of a day D As early as 1870 S Newcomb suspected the latter but it was not until 1939 that S W Brown definitely attributed the discordances to changes in speed of rotation of the Earth

Time The unit of time defined by the rotation of Earth about its axis is the second of mean solar time which is defined as 1/86 400 of a mean solar day On account of variations in speed of rotation this is not a constant unit of time as judged by a perfect clock Within the past 200 years changes as great as 5 parts in 10^8 have occurred In 1956 the International Bureau of Weights and Measures redefined the unit of time so as to make it identical with the second of ephemeris time which is a constant unit of time Ephemeris time is defined by the orbital motion of the Earth about the Sun The measure of ephemeris time was adopted so that it is approximately the same as the average of mean solar time over the interval 1700–1900

In measuring variations in speed of rotation it is desirable to use the second of ephemeris time as a standard However in the past all observations were referred to the second of mean solar time which is a variable unit

Tidal friction The Moon raises tides in the ocean Friction carries the maximum tide ahead of the line joining the center of the Earth and Moon (Fig 1) The resulting couple diminishes the speed of rotation of the Earth The couple reacts on the Moon to increase its orbital momentum The sum of the angular momentum of the Earth and the orbital momentum of the Moon remains constant The effect is to increase the size of the orbit of the Moon and to diminish its angular speed about the Earth

With respect to ephemeris time the effect of tidal friction is to increase both D and the lunar month Because of the change in the lunar month the Moon has an orbital deceleration in terms of ephemeris time The proportional change in D however is greater than in the month Hence in terms of mean solar time the Moon appears to have a secular orbital acceleration an effect disc

by E Halley in 1693 from a study of ancient eclipses If the Earth has changed its speed of rotation since ancient times the path of an eclipse which occurred 2000 years ago would be displaced in longitude with respect to the path that would have occurred if the speed had remained constant

The secular acceleration when discovered was puzzling it was not then attributed to tidal friction In 1786 P S Laplace announced that the secular acceleration was due to a neglected effect in the gravitational theory of the motion of the Moon However in 1853 J C Adams found that gravitational theory could account for only about half the acceleration found by Halley It has become clear since then that the rotation of the Earth is slowing down because of tidal friction

Researches by S Newcomb W de Sitter E W Brown Sir Harold Spencer Jones and others have shown that the Earth has irregular variations in addition to the secular retardation The orbital motions of Earth Venus and Mercury about the Sun show discordances similar to that shown by the Moon but proportional to their mean speeds

Energy dissipation The energy lost by tidal friction must be dissipated somehow G I Taylor and Sir Harold Jeffreys have studied the effects in shallow waters Jeffreys considered that the Bering Sea

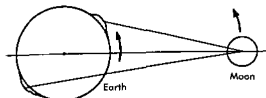


Fig 1 Couple produced by tidal friction

indicate that only about 10% of the required effect can be accounted for by tidal currents in shallow seas This question is therefore still open

Periodic variations Seasonal variations were detected in 1937 by N Stoyko who used a number of pendulum and quartz crystal clocks The early determinations of the seasonal variation appear to be too large judged by results obtained with improved quartz crystal clocks since 1950 and with atomic clocks since 1956 by a factor of two The Earth is behind its mean position by about 0ˢ035 at the end of May and about 0ˢ030 ahead at the beginning of October

The seasonal variation is probably caused by an exchange of momentum between winds and the crust of the Earth

Lunar tidal variations of periods 27 55 days and 13 66 days were predicted by Jeffreys in 1928 The amplitude of each term is about 0ˢ001 in time These terms were detected in 1955 from observations made with the Washington D C and Richmond Florida photographic zenith tubes of th Naval Observatory [W

REVOLUTION ALONG ORBIT

Motion of Earth about Sun is seen as an apparent annual motion of the Sun along the ecliptic That the effect is caused by motion of Earth and not of the Sun is proved by the annual parallactic displacement of near stars and by the aberration of light causing an apparent annual displacement of all stars on the celestial sphere See ABERRATION OF LIGHT

Period of revolution The true period of revolution of the Earth around the Sun is determined by the time interval between successive returns of the Sun to the direction of the same star This interval is the sidereal year T = 365 days 6 hours 9 min 9 5 sec of mean solar time or 365 25636 mean solar days The period between successive returns to the moving vernal equinox is the tropical year T = 365 days 5 hours 48 min 46 0 sec or 365 24220 days Chronology is based on the tropical year (see CALENDAR) The period between successive passages at perihelion is called the anomalistic year T'' = 365 days 6 hours 13 min 53 0 sec or 365 25964 days

Mean radius of orbit The mean distance from Earth to Sun or semimajor axis of the Earth s orbit is the astronomical unit of distances in the solar system Its absolute value fixes the scale of the solar system and the whole universe in terms of terrestrial standards of length This value can be determined by a variety of methods the results are usually expressed in terms of solar parallax or more precisely the Sun s mean equatorial horizontal parallax p which is the apparent diameter of the equatorial radius r of the Earth at the mean distance a of the Sun The relation between r and p is $a = r \sin p = 206\,265r/p$ if p is in seconds of arc The equatorial radius r of the Earth is 6 378 388 meters

To measure the solar parallax geometrical gravitational and ph

Geometr

volve the d

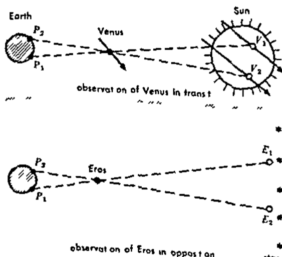


Fig 2 Determination of solar parallax

tion of the parallax of a nearby planet (Mars Venus) or asteroid (Eros) at its closest approach to Earth (opposition or transit) Because the relative distances in the solar system are accurately known in terms of the astronomical unit the absolute measurement of one distance gives the scale of the system (Fig 2)

The most accurate trigonometric determination, derived by H Spencer Jones in 1941 from an international campaign of observation of the minor planet Eros near its opposition in 1930–1931 is $p = 8''790 \pm 0''001$ corresponding to a mean solar distance $a = 149\,675\,000 \pm 17\,000$ km ≅ 93 004 000 ± 11 000 mi This determination is now known to be in error by about 0 1%

Gravitational methods The gravitational methods involve the determination of the ratio of the mass m of the Earth to the mass M of the Sun from the perturbations in the motion of a minor planet (Eros) caused by the Earth The method rests essentially on a comparison between length $l = gt^2/4\pi^2$ of a pendulum of double oscillation period t (seconds of mean solar time) on Earth which gives the acceleration of gravity $g = Gm/r$ in terms of gravitational constant G and the length of the radius a of the Earth's orbit related by Newton's law to the duration T of the sidereal year (in seconds of mean solar time) by

$$4\pi a^3/T^2 = G(M+m) = g(1+M/m)$$

the Earth s radius is not needed to compute a but it is necessary to make allowance for the mass of the Moon relative to the mass of Earth

This is at present the most accurate method and the best result derived by W Rabe in 1948 from 50 years of observations of Eros at its nearest approaches to the Earth is $p = 8''7984 \pm 0''0004$, corresponding to a mean solar distance $a = 149\,532\,000 \pm 7\,000$ km ≅ 92 915 000 ± 4 500 mi

Another method is to measure the echo time of radar signals reflected by Venus Provisional determinations in 1958 and 1959 correspond to a solar parallax $p = 8''8022 \pm 0''0001$

Physical methods The physical methods rest on a determination of the ratio of the mean orbital velocity $V = 2\pi a/T$ to the accurately known velocity of light This ratio can be derived either from the annual variations of the radial velocities of ecliptic stars (or occasionally of planets) determined by observations of the Doppler shift of spectral lines or with less accuracy from the constant of aberration

Orbital velocity The mean orbital velocity of the Earth is 29 80 km/sec ≅ 18 5 mi/sec The curvature of the orbit as measured by the departure from its tangent at the end of the arc traveled in 1 sec of time is 0 296 cm ≅ 0 117 in , the acceleration toward the Sun is numerically twice this quantity or 0 593 cm/sec², in other words the attraction of the Sun at the mean distance of the

Earth GM/a^2 is about 0.605% of the attraction of the Earth on bodies at its surface, Gm/r^2. To produce this attraction at a distance equal to a/r = 23,444 Earth radii, the mass of the Sun is therefore approximately 0.605 $(23,444)^2$ = 333 000 times the mass of the Earth.

Eccentricity of orbit. The eccentricity of the Earth's orbit was initially determined by the variations of the apparent diameter of the Sun's disk; it is accurately determined by the variable speed of the Sun's apparent motion along the ecliptic and the laws of elliptic motion (*see* PLANET). The nonuniformity of the Sun's motion manifests itself in the equation of time, which is the difference between apparent (or true) and mean solar time. This difference arises in part from the obliquity of the ecliptic and in part from the eccentricity of the Earth's orbit. The present value of the eccentricity of the Earth's orbit is 0.01675; it decreases slowly under the effect of planetary perturbations, the present rate of decrease being −0.000042 per century counting from 1900. However, the eccentricity will not decrease to zero and will increase again after reaching a minimum. The Earth is at perihelion on January 2 and at aphelion on July 2; the corresponding variation of temperature tends to reduce the seasonal amplitude in the Northern Hemisphere and to increase it in the Southern Hemisphere. [C.D.V.]

Earth deformations and vibrations

Alterations of form or shape of the earth. These changes and vibrations range from minute to great in magnitude at intervals differing from regular to irregular and from extremely short to long portions of the geologic history of the earth. Large displacements of rock masses along a geologic fault or the shock of an earthquake can be impressive to a casual observer, but many types of deformations are revealed only by instruments of the highest sensitivity, or by precise measurements extending over tens or hundreds of years. A systematic survey of these and related questions is briefly developed in this article and may be carried further by consulting the articles cited by cross reference.

Background considerations. The earth is an engine powered by heat and at times by gravitational energy which has been operating vigorously for some 4 000 000 000 years. Some of the results to date of this activity may be seen in the present surface features of the earth, for example, the distribution of continents, oceans, and mountain systems. More basic to the understanding of the energetics of the evolving earth is the concealed structure of its interior, as it appears revealed by the behavior of seismic waves. These seem to indicate a fluid core, thick mantle, and thin superficial crust above a well defined Mohorovičic discontinuity. The components of the earth are changing features of a dynamic environment. It is well known that the continents were not always as seen today. It is even possible that they did not always occupy the same relative geographic positions (*see* ROCK MAGNETISM). Today the fluid core is considered as having a diameter about 55% of the earth's diameter (*see* EARTH INTERIOR). The condition of the core 1 000 000 000 years ago may have been very different. The grand objective of the earth sciences is to recapture the full history of the earth and to understand its significant events.

Systematic geophysical studies. Much of the evidence concerning the present nature and past evolution of the earth is obtained from observation of its motions and deformations. The kinds of motions available for study are conveniently classified into three groups in respect to their time scales: (1) major events of geological history, (2) dynamical behavior patterns of the earth, and (3) elastic waves in the earth.

Major events of geological history. These include major events throughout a long time scale of several billion years. Investigations of these events focus on the growth of continents, oceans, and mountains, and on the transport of large amounts of material by lava flows from depth. These subjects are the province of special geophysical studies (*see* OROGENY; TECTONOPHYSICS; VOLCANOLOGY). The abundant evidence of large horizontal shearing displacements along extensive fault systems seems to be especially significant. The San Andreas fault, extending northwest in a nearly straight line for 900 miles across California, shows evidence of accumulative clockwise shear

ments of 250 km along Pioneer Ridge (38°30′N between 127° and 136°W) along an east-west fault. These are examples of a significant type of mobility of the crust. They represent large continuing shear strains and associated shearing couples. Shearing couples on this scale seem difficult to explain without postulating a mobile subcrust supporting slow convection on a large scale.

Dynamic behavior patterns of the earth. During the span of historic time observations of astronomy and geodesy have established significant features of the earth's dynamic behavior. The rate of precession of the equinoxes provides the best value of the ratio H involving the moments of inertia of the earth, $H = C - [(A + B)/2C]$, where A, B, and C are principal moments of inertia. The period of the 14-month Chandlerian wobble provides a value for the Love number k. The rate of transfer of angular momentum from the earth to the moon and sun by virtue of tidal torques has been best determined by astronomical observations during the last several hundred years. Observations of earth ... meters, horizontal pendulums, and

(which is slightly less than 1 hour). Accordingly, the solid earth probably responds to tidal forces in an essentially static manner. Unfortunately, the modifications in this response caused by ocean tides are still undetermined. *See* EARTH TIDES; *see also* GEODESY; TERRESTRIAL GRAV

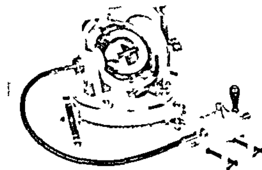
Elastic waves in the earth Such waves traversing the earth from earthquake foci or from man made explosions have furnished all the precise data about the earth's internal geometry and of the changing values of its elastic parameters with depth The recent large atomic explosions whose time and place of origin are accurately known have significantly improved precision of seismic interpretations concerning the earth's interior Furthermore the increasing availability of large digital computers has made feasible the heavy computational programs involved in power spectra analyses of seismic records and of earth tide records Such analyses increase the amount of significant information deducible from the records See SEISMOLOGY [L B SL]

Bibliography J C Crowell The San Andreas fault in southern California *Proc 21st Intern Geologic Congress Norden Germany* 1960 H Jeffreys *The Earth Its Origin History and Physical Constitution* 4th ed 1959 W H Munk and G F MacDonald *The Rotation of the Earth A Geophysical Study* 1960 V Vacquier Measurement of horizontal displacement along faults in the ocean floor *Nature* 183(4659) 452–453 1959

Earth inductor

An instrument for measuring the dip angle of the earth's magnetic field Essentially it is a coil of wire designed to be rotated about an axis parallel to the plane of the coil As the coil turns in the earth's field

ented about a vertical axis so that the rotational axis is approximately (within a few minutes of arc) in the magnetic meridian In that position the deviation of the rotational axis from the inclination of the magnetic field produces upon rotation of the coil a voltage readily detectable with a galvanometer Electrical connection is made between the coil and the galvanometer through a two segment commutator The coil frame is mounted to turn upon conical bearings of agate and is equipped with a vertical circle reading to about 10' of arc in the better instruments Errors due to mechanical maladjustments of the coil axis are made negligible by taking a series of readings for plus and minus rotations of the coil vertical circle east and west and commutator up and down Readings of the circle for axis inclined are compared with readings for axis vertical (adjusted with the coil at rest) as indicated by a highly sensitive level bubble mounted within the frame of the coil itself



Observatory earth inductor (U S Coast and Geodetic Survey)

The galvanometer used with the earth inductor is very sensitive but has little distorting effect on the earth's field at distances of half a meter It consists of two or more small astatically balanced permanent magnets suspended with a reflecting mirror from a fine quartz fiber Fixed coils of many turns mounted near the suspended magnets carry the current that must be measured or detected

The accuracy of a fine quality observatory earth inductor in good adjustment is of the order of 0 1' of arc while the smaller, lighter models for use on tripods in field work yield results to perhaps 0 5 The earth inductor is an absolute instrument in that theoretically it does not require calibration by comparison with an accepted standard In practice such comparison is always made and an index correction is adopted The deviation from standard usually less than a minute may be caused by slight traces of magnetic impurity in the frame or other stationary parts of the instrument or by other factors which cannot be completely defined and eliminated

The earth inductor has almost completely supplanted the dip circle throughout the world for precise measurements of magnetic inclination [J H NE]

Bibliography See MAGNETOMETER

Earth interior

Features of the interior of the earth are revealed through the application of the principles and techniques of physics and related sciences and subsequent mathematical analysis Although the earth's interior is not available to direct visual observation many of its properties are nevertheless firmly established This article discusses the following properties spherical stratification mean density and moment of inertia seismic data density variation pressure variation acceleration due to gravity incompressibility rigidity composition internal heat and temperature and magnetism

Spherical stratification The earth is a planet of mean radius 6371 km The composition of the outermost 35–40 km depends upon the geographical region differences being specially marked between continental and oceanic regions The differences rapidly become unimportant as the depth increases and for most of the interior the chief deviations from spherical symmetry are due to the internal oblatenesses (flatness at axial poles) of internal surfaces of constant density and composition The oblateness diminishes with increasing depth and the ratio of the surface oblateness measured by the ratio of

the difference between equatorial and polar radii to the mean radius is only about one three hundredth For these reasons the earth may to good approximation be treated as concentrically layered the chief properties varying only with the depth

The exceptional outermost 35 km include the outer layers sometimes called the crust which is bounded by the Mohorovicic discontinuity so called because of the work of the Balkan seismologist Mohorovicic in 1909 Under continents the discontinuity is about 40 km deep it is somewhat deeper under mountain chains but lies only about 5 to 10 km below ocean floors

Mean density and moment of inertia For the earth compared with water mean density is 5 517 Certain astronomical data measurements of the earth s shape and the dynamics of the earth moon system show that the earth s moment of inertia is 0 83 times that of a body of constant density with the same size and mass The smaller the moment of inertia of a body the greater the degree of central concentration in it These results show that the earth s deep interior is much denser than rocks found near the surface and provide two important numerical criteria on the internal distribution of matter

Seismic data Evidence from seismology leads to much further detail A large earthquake sends out seismic waves which penetrate the whole interior of the earth The waves rise again to the surface and are recorded on seismographs in the thousand or so seismological observatories that have been set up in many countries Through solid parts of the earth both *P* (primary) and *S* (secondary) waves are transmitted *S* waves having about two-thirds the speed of *P* waves Through fluid zones only *P* waves are transmitted From the times of arrival of the various wave pulses at the earth s surface seismologists have deduced values of the *P* and *S* velocities throughout much of the interior See SEISMOGRAPH SEISMOLOGY

The results supply the chief evidence on the nearly spherical character of the earth s subcrustal stratification The variation of the velocities with depth then enables the whole interior to be charted out into regions whose boundaries are characterized usually by sharp changes either in the *P* or *S* velocity or in the rate of change with respect to depth The following table based on work of

Table of earth s layering properties

Region	Name	Range of depth miles	*P* velocity km/sec
A	(Crust)	0–35	Very variable
B C	Upper mantle	35 1000	8–9
D	Lower mantle	1000–2700	9–13
D		2700–2900	
E	Outer core	2900–5000	8 10½
F	Transition region	5000–5100	Uncertain
G	Inner core	5100–6371	11

H Jeffreys and K F Bullen summarizes in broad terms the earth s layering as inferred in this way

The separation into mantle and central core was revealed by R D Oldham in 1906 through the presence of a partial shadow in observations of the main *P* waves reaching the surface from a distant earthquake The waves which cross the boundary from mantle to core are refracted toward the earth s center leaving a shadow between angular distances of 105 and 142° from the earthquake source The boundary between mantle and core is sharply defined In 1913 B Gutenberg calculated its depth to be 2900 km (1800 miles) Jeffreys has shown this figure to be accurate within about 4 km

The transition between upper and lower mantle is gradual the depth of the separating surface being taken somewhat conventionally

The separation into outer and inner core was revealed by I Lehmann in 1936 through the refraction away from the earth s center of certain descending *P* waves which enter the inner core

Density variation At any point inside the earth the values of the *P* and *S* seismic velocities depend principally on the local density incompressibility and rigidity The two latter moduli describe the elastic behavior incompressibility or bulk modulus measures resistance to change of density under pressure rigidity measures resistance to distortion of shape A fluid has negligible rigidity and for this reason does not transmit *S* waves

The *P* and *S* velocities are fairly well known down to the bottom of the outer core and so provide two equations toward determining the density and elastic moduli *S* waves are observed only in the mantle but the *S* velocity in the outer core can be taken as zero because of its fluidity as discussed later

The additional evidence needed to determine the actual values of such properties as density and rigidity in the earth includes use of the known mean density and moment of inertia of the earth and of the theory of gravitational attraction Seismic data are further used because density changes due to the increasing pressure in the earth are linked with the incompressibility Finally the matching of laboratory experiments on rocks against *P* and *S* velocity data yields a useful partial correlation between velocities and densities

From this evidence Bullen has calculated that the density increases from an assumed starting value of 3 3 just below the crust to 5¾ at the base of the mantle jumps abruptly to 9¼ at the top of the outer core and reaches 11½ at a depth of 5000 km (3100 miles) The density in the inner core is less definitely determined but most probably lies between 14½ and 18 The assumed starting value of 3 3 rests on geological and related laboratory evidence any correction which this value may need can have little effect on the values below the upper mantle

Pressure variation In the earth values of the pressure are derived from the same series of calcu

lations The greatest pressure realized in high pressure laboratory experiments about 300 000 atmospheres (atm), occurs about 800 km below the earth s surface The pressure is 1 400 000 atm at the bottom of the mantle and is 3 500 000–4 000 000 atm at the center of the earth

Acceleration due to gravity This acceleration *g* remains within 1% of 990 cm/sec (32 4 ft/sec) down to a depth of 2400 km (1500 miles) rises to a maximum of about 1040 cm/sec at the bottom of the mantle and then falls steadily inside the core to zero at the center

Incompressibility On the whole incompressibility increases steadily with depth Below the outer mantle its variation with pressure is remarkably smooth and little affected by possible variations of chemical composition The inner core is about 10 times as incompressible as ordinary steel

Rigidity The mantle is solid throughout (apart from the oceans and limited pockets of volcanic matter) since *S* as well as *P* seismic waves are everywhere transmitted The rigidity in fact increases steadily with depth in the mantle which at the bottom is 4 times as rigid as ordinary steel The term rigidity here relates to the behavior of a material under rapidly changing stresses as during the passage of earthquake waves Convection currents or other forms of solid flow taking place over geologically long periods of time are not precluded Whether such flow actually occurs is controversial The outer core is very much less rigid than the mantle so that the state of the outer core is almost fluid or molten Negative evidence of this is the failure to detect *S* waves below the mantle The positive evidence lies in direct calculations by H Takeuchi and M S Molodenski of the core rigidity based on the known rigidity of the mantle and on measurements of its tidal straining *See* EARTH TIDE

The inner core is probably solid The refraction of *P* waves at its boundary implies a jump in the *P* velocity from outer to inner core Such a jump requires a jump in either the incompressibility or the rigidity and the smooth variation of incompressibility with pressure elsewhere in the earth makes the latter much the more probable Bullen estimates the rigidity of the inner core to be between 2 and 4 times that of ordinary steel Calculations in theoretical physics on the behavior of matter at very great pressures support this view

Composition Assuming the immediate subcrustal density to be 3 3 calculations show that there must be a marked change of chemical composition or else a physical transformation of some kind brought about by the pressure somewhere in the [illegible] divi[illegible] [illegible] may be much greater The part *D'* of the lower mantle appears to be of fairly uniform composition while in *D'*, the lowest 200 km there is evidence of some small accumulation of denser materials There is no sharp boundary between *D'* and *D''*

If the immediate subcrustal density were as high as 3 7, the whole mantle below the crust would be nearly homogeneous, apart from the effects of compression

There is general agreement that the outer mantle consists of ultrabasic rock A common view is that the mineral olivine consisting of iron magnesium silicate predominates in the region *B* although there are other possibilities The region *C* may be either a transition region in which the high pressure brings about a change of crystal form or a region in which the chemical composition changes gradually The region *D'* probably contains the same chemical ingredients as *B* and *C* but according to F Birch in the form of distinct phases such as silica magnesia and iron oxide The mantle may also contain an appreciable quantity of uncombined iron There is as yet no clear view on the composition of such denser materials as may have accumulated in *D'*

Chemical analysis of meteorites which are possibly samples of the interior of an earthlike planet and knowledge of the earth s high core density led to the long standing theory that the core consists predominantly of molten iron probably mixed with nickel In 1948 W H Ramsey suggested instead that the outer core consists of the same chemical ingredients as the lower mantle but in a high density metallic modification caused by the huge pressure On Ramsey's theory the Earth Mars and Venus could be of the same over all chemical composition whereas this is not possible if the earth's mantle and outer core are chemically distinct It has not yet proved possible to discriminate between the two theories although the older is still slightly favored Evidence from theoretical physics suggests that the representative atomic number for the outer core is somewhat less than for iron and nickel, but alloying of iron with some other material might account for this independently of Ramsey's theory

The inner core probably does consist largely of iron and nickel and there is also some suggestion of a gradually changing composition The region *F* is probably a transition region but the seismic data are as yet too indefinite to indicate its nature

Internal heat and temperature The earth is commonly assumed to have been formed molten H Jeffreys has shown that the mantle would then have solidified from the bottom upward He attributes the formation of mountains to thermal contraction largely confined to the outermost 700 km a region in which incidentally all recorded earthquakes have originated In 1952 from considerations of theoretical chemistry H Urey suggested that the earth was accumulated from small objects at temperatures "perhaps near zero and certainly less than a few hundred degrees centigrade" though the gravitational energy of accumulation could have produced higher temperatures

The fact that the earth has not by now cooled to zero temperature is due to the presence of radio activity below the surface The solidity of the mantle sets an upper limit to the total quantity of radioactivity and shows the latter to be largely confined to the crust Until lately radioactivity was thought to be confined largely to continental areas but measurements of heat flow over the oceans taken since 1953 show that this is not so Small radioactivity is likely to occur also below the crust but its detailed distribution is uncertain *See* EARTH (HEAT FLOW)

In an extension of the work of Jeffreys J A Jacobs suggested in 1953 that the earth started solidifying from the center outward about the time that the mantle started solidifying This theory assumes the whole core to be of the same chemical composition different from that of the mantle and attributes the division into molten outer and solid inner core to pressure and temperature gradients in the core

From the fact that the whole mantle is solid, R J Uffen concluded in 1954 that the temperature nowhere exceeds 5000°C some others would put the figure somewhat higher J Verhoogen in 1954 showed that the temperature at the bottom must be at least 2000°C These figures set approximate limits to the lower mantle temperature The temperature at the earth's center may exceed the temperature at the base of the mantle by about 500° or less

Magnetism Early theories attributed the earth's magnetism to permanently magnetized iron in the deep interior This is now considered improbable because of the seismic and other evidence on the fluidity of the outer core Nevertheless it is now well established that the earth s magnetic field does originate predominantly in the deep interior The most widely accepted current theory the dynamo theory of Elasser and Bullard attributes the earth s magnetism to currents flowing inside the fluid outer core and involves the conversion of mechanical energy into electromagnetic energy The theory is compatible with either a molten iron outer core or with the Ramsey theory which requires the outer core even if composed of the same ingredients as rocks to have the electrically conducting properties of a metal It is also compatible with the presence of a solid inner core the boundary of which would enable convection currents of the required type to occur in the outer core The theory requires a source of energy for the convection currents Radioactivity has been suggested but is thought by many to be inadequate This is the chief difficulty the theory has to face

See EARTH TERRESTRIAL MAGNETISM [K E BU]

Bibliography D R Bates (ed) *The Planet Earth* 1957 K E Bullen *Introduction to the Theory of Seismology* 2d ed 1954 K E Bullen *Seismology* 1954 B Gutenberg *Physics of the Earth's Interior* 1959 H Jeffreys *The Earth Its Origin History and Physical Constitution* 4th ed, 1959

Earth resource patterns

The physical character and distribution of natural resources at the face of the earth No section of the earth is exactly like any other in its resource endowment Nevertheless latitudinal differences in insolation the great difference between land and ocean and geological composition of the earth s crust together provide the basis for distinguishing definite geographical patterns of resource availability over the world

Delineation of the earth s resource pattern begins with differentiation between land and marine resources Although marine resources have been used by men since earliest times the 2 700 000 000 people on the earth in the mid twentieth century are highly dependent upon the resources of the land for their continued existence

Five principal resources associated with land are fresh water agricultural soils mineral deposits forest lands, and grasslands Such other resources as the native animal life and genetic stocks of plants and animals are very mobile and currently cannot be considered significant in differentiation of resource patterns On the other hand the five principal resources do have unique associations which differ from one broad area to another

The underlying causes for distinctive features in the pattern of resource endowment are the regime of energy receipts from the sun the effects of planetary atmospheric circulation on the distribution of moisture and heat the geological composition of the continents and islands and the structural history of different sections of the earth's surface

Ready keys for understanding the resource pattern of a section of the earth s surface are given in the character of climate earth surface configuration and rock composition A basic pattern is outlined in regional climates with associated characteristic types of agricultural forest and grazing lands and water availability On this is placed an overlay of differences in rock composition which alters the pattern within the climatic regions A second overlay of differences in surface configuration produces further alteration on the previously shown patterns

Eleven regional climatic types (numbered consecutively on map) in four groups are recognized in describing the earth s resource pattern This number of types is fewer than that normally employed to describe regional climates but is considered adequate to outline the basic resource pattern Distinction is made between the so-called humid climates and the water deficient climates with subtypes as follows

Humid microthermal
- (1) Polar and icecap
- (2) Tundra
- (3) Taiga
- (4) Puna

Humid mesothermal
- (5) Upper midlatitude
- (6) Humid subtropical

Humid macrothermal
(7) Wet and dry
(8) Rainforest
Water deficient
(9) Desert
(10) Semiarid
(11) mediterranean

Humid microthermal regions These areas of predominantly low temperatures are characterized by unfavorability to crop and natural vegetative growth and under present techniques by a relatively low carrying capacity for people

(1) In polar and icecap areas available resources are dominantly marine and land animal life on which the very sparse human settlement is almost wholly dependent Despite the enormous icecap area of the Antarctic settled sections are exclusively in the Arctic Polar regions may be of future importance for mineral deposits but their geology is as yet relatively unknown

(2) Tundra except for minor alpine locations is entirely within the northern hemisphere The principal resources are lichens and the native animal life like reindeer and caribou which can make use of these as food Parts of the tundra may be considered a grazing land as managed herds of reindeer are pastured nomadically Potential mineral deposits are imperfectly known although a few commercial workings are now located within tundra regions

(3) M[illegible]
norther[illegible]
are car[illegible]
the most valuable known resource of the microthermal regions Varieties of spruce fir and larch are of particular significance to the pulping industries Some taiga may be considered potential agricultural land but the possible incidence of frost in every month and the very short growing season (80 days or less) do not presage major agricultural development under present techniques Mineral deposits of uranium copper iron and others are known and exploited in the taiga It is also a commercial source of fur bearing animals

(4) The puna type of climate is found where high elevations above sea level result in temperatures which place certain plateaus, such as the high plateaus of Tibet and limited sections of the South American upland in this group In these treeless places the principal resource is low productivity grazing land Cultivated lands are limited to hardy small grains and root crops South American puna areas contain some significant metal deposits particularly copper and tin

Humid mesothermal regions On the average temperatures are intermediate in these parts Considered in the light of present day technology the heart of the world s natural resource base is in the mesothermal regions They contain a large share of the crop and pasture lands and some productive [illegible]

(5) Upper midlatitude climate contains most of the lands adapted to the raising of wheat barley and rye Certain areas such as the North American Corn Belt and the Danubian Valley also produce maize Extensive forage cropping supports dairying and meat animal raising Central North America western and central Europe, the nonarid part of southern Soviet Union and Argentina account for most of the area within this climate

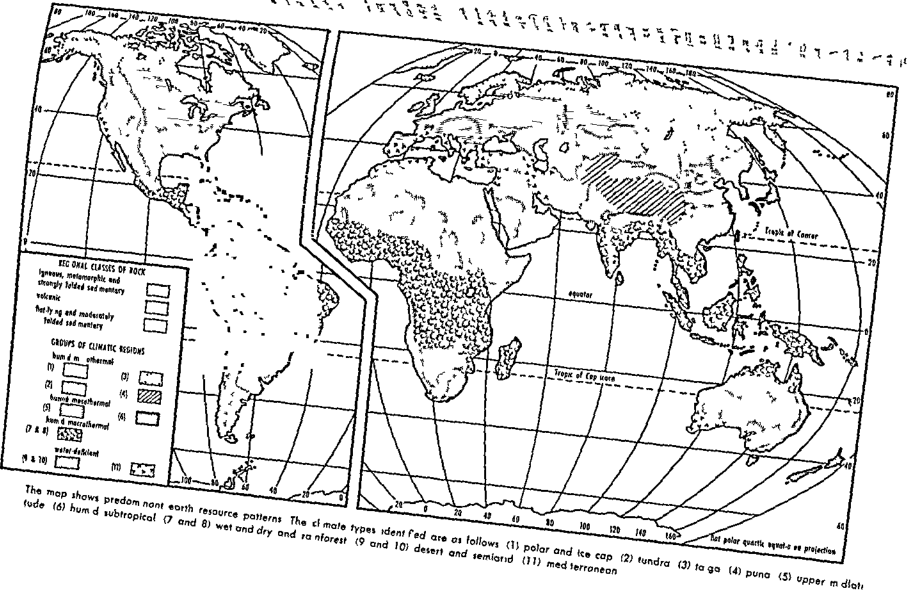
(6) Humid subtropical areas have an ample water supply and relatively long growing season (200 days or longer), making these rice and cotton lands highly productive Soils in these areas unless well fertilized are less capable of sustaining cultivation over extended periods than those of the poleward —mostly northern—parts of the middle latitudes

Humid macrothermal regions Predominantly winterless regions of warm to hot temperatures such areas are divided according to the regime of rainfall (7) wet and dry, with a pronounced dry season and (8) rainforest, with year around precipitation Macrothermal regions have year round growing seasons and lateritic soils Problems induced by fungal growth bacterial disease and insect abundance handicap the agricultural land resources Within the great alluvial valleys extensive flooding also may be disrupting Agricultural land resources are developed only in spots Particularly within the rainforest regions extensive and rapid growing forests occur but are mostly unexploited commercially in the twentieth century economy Sites of enormous potential hydroelectric generation remain largely undeveloped

Water deficient regions Receipts of moisture are scant or lacking during much of the year Except where exotic water supplies are available for irrigation the agricultural lands inherently are less productive than those of any humid region with similar land surface There are three major subdivisions

(9) Deserts differ strikingly in their form and in temperature conditions but everywhere present meager resources for agriculture Where water is available desert oases blossom but vast areas contain only scrub growth ephemerals or virtually no vegetation Some grazing resources are available but carrying capacity of deserts is very low Sparsity of vegetation has made mineral prospecting and exploration somewhat easier than in vegetation covered areas and in this century desert occupance often started with mineral discoveries

(10) The semiarid regions are basically grasslands which have grazing as their characteristic resource Because of cyclic rainfall variability men have converted the inherently fertile soils as in China and the United States to cereal growing during periods of higher rainfall Rainfall fluctuations however make the land resource unstable under cultivation For this reason these regions have suffered consistently from the accelerated and destructive erosion resulting from unsuited cultivation and overgrazing Semiarid lands are responsive to and are most productive under irriga



The map shows predom nant earth resource patterns The cl mate types ident fed are as follows (1) polar and ice cap (2) tundra (3) ta ga (4) puna (5) upper m dlatitude (6) hum d subtropical (7 and 8) wet and dry and ra nforest (9 and 10) desert and semiarid (11) med terranean

tion Irrigation is not developed to its full potential because its value depends heavily upon temperature conditions and location with reference to markets

(11) Regions of mediterranean climate because of their winter rainfall generally are classed as humid lands However the greater part of the growing season is water deficient and the most productive agricultural lands are dependent on irrigation Agricultural lands are the major resource since water deficiency is pronounced enough to discourage forest productivity Major mineral deposits may complement agricultural lands in a few areas

Rock composition and surface configuration Imposed on the basic land pattern induced by climatic differences there are variations in rock composition and surface configuration which cause intraregional differences within the pattern Although not exactly the same in their effects the variations caused by these two geographical elements are often concomitant and may be treated together as follows

1 Rock composition and structure
   - *a* Flat lying and moderately folded sedimentary rocks
   - *b* Igneous metamorphic strongly folded sedimentary rocks
   - *c* Volcanic rocks
2 Surface configuration
   - *a* Flood plains and other flat or gently sloped surface
   - *b* Mountains and maturely dissected hill lands plateau faces or faces of cuestas

These elements of crustal variation produce the following six geographical differences in resource endowment

First all major agricultural lands are on flat lying or moderately folded sediments and have gentle slopes well exemplified in such alluvial valleys as the Mississippi Nile Huang and Ganges Brahmaputra

Second productive secondary agricultural land resources are located on volcanic areas where soils have been formed through weathering or wind action The Deccan section of India for example is one of the largest areas of this kind

Third agricultural lands are extremely limited on igneous rock areas *no matter what the surface* configuration particularly *in regions* north of 40°N latitude as *illustrated* in the occupance of the Laurentian shield area (Canada) *and the* Fenno Scandian shield (Europe) Exceptions are found in the humid tropical and subtropical *climates* where weathering has proceeded long enough to produce a substantial soil mantle as *on the* Piedmont of [illegible]

Fifth mountainous areas are *important* catchment areas and sources of *fresh water* and the services which may be derived from water Most hydroelectric generation, or generation potential, is associated with mountains In arid and semiarid regions mountains are sources of water for irrigation, domestic and industrial water supply wood products and warm season grazing lands

Sixth mineral resources have definite patterns which are associated with rock structure and composition Major deposits of coal petroleum natural gas and lignite are with few exceptions found in flat lying or gently folded and faulted sedimentary rocks as in Texas oil fields and the coal fields of the Allegheny Plateau Sedimentary nonmetallics (the phosphates sulfur, nitrates and limestone) as well as bauxite and uranium (carnotite) also are associated with sedimentary rocks

Associated with the igneous and metamorphic rock areas are most metals for example iron lead, copper tin the ferroalloys gold and silver Most gems and some nonmetallic minerals (mica asbestos) are found in the same association Uranium (pitchblende) occurs in these rocks

While these associations are well *recognized the* mineral deposits have an erratic geographical occurrence

Employed and potential resources Resources have meaning insofar as they are *placed* in use or are available for future exploitation Distinction must be made between the employed and the potential resources In practice, this *distinction* is complex but here only the simple geographical distinction will be noted Employed resources are those which are significant to the present support of mankind at least locally In general the *denser* the population and the more advanced the *technical* arts of an area the greater the need for *production* from resources and employed *resources* become more nearly synonymous with all known *resources* Thus the recognized resources of the European peninsula and of northeastern United *States are* mainly employed resources On the other *hand the* natural resources of the Amazon basin *or* of Alaska and western Siberia are still largely *potential resources*

Marine resources Although the physical and biotic geography of the oceans is much less *fully* explored than that of the *continents*, enough is known to indicate that both living and mineral resources extend far beyond *those presently* exploited The employed resources are rather *sharply* localized The principal exploitation of *marine* animals and vegetation is (1) over the continental shelves (2) in the vicinity of the mixing of warm and cold currents, (3) near large *upwellings* which characteristically occur in *lower* middle latitudes and (4) adjacent to densely settled *countries* Thus the North Atlantic near Europe and from New England to Newfoundland contains *heavily* exploited fishing grounds as do the *seas near* Japan Korea and southern California and along the waters of the Gulf of Mexico *Minerals of the seas are* employed resources in few localities, salt is the principal mineral extracted mainly in *dry* low latitude areas

There seem to be large potential marine resources which include (1) the population of life forms now exploited in some parts of the world but not in others, (2) animal and vegetative species currently unused, (3) fresh water from desalted sea water, (4) the minerals which are in solution are precipitated to the ocean bottom or lie within rock below the water One of the interesting speculative resources appears in the large quantities of so called manganese nodules that cover some sections of sea bottom at intermediate depths Several of the minerals not commonly found in land deposits like manganese and nickel occur in these deposits *See* MARINE RESOURCES

Resources of the continents The resource pattern of the earth may be summarized in a brief description of that for each continent and its neighboring waters

Eurasian continent As the largest land mass in the world the Eurasian continent has the largest area of agricultural land in use a very extensive total forest land area and a wide variety of mineral deposits Great differences mark the several sections of the continent The most productive agricultural areas are generally near the edge of the land mass in western and central Europe European U S S R India, and mainland China Much of interior Siberia central Asia and Asia Minor shut off from productive agriculture by cold and drought contain forest resources of major potential significance and mineral reserves including petroleum coal and the metals The southeastern and eastern borders of the heartland moreover have some of the great but still undeveloped hydro generation sites of the world Off the coasts of west Europe north China and south Siberia are the two most productively employed fisheries of the world

Africa and Australia Africa is handicapped by heat and drought A major section of the continent must be classed as desert or semiarid with few exotic water sources Much of the remainder has wet and dry or rainforest climates with the attendant handicaps to forest or agricultural exploitation The east African highlands from Ethiopia southward the Cape area of South Africa and the Nile Valley are the only noteworthy exceptions Except in the Nile Valley there are still potential agricultural land resources but they are comparatively minor Africa's chief resources over the long term may prove to be mineral including fuels and other energy resources Although still far from fully explored major metal deposits include copper bauxite, and iron and diamonds gold and uranium are well known Large potential but undeveloped hydroelectrical resources also exist

Similar general remarks may be made about Australia whose smaller area is covered mostly by desert semiarid and tropical wet and-dry environments Most productivity and population are peripheral especially in the southeast

South America The land resource is dominated by the unbroken extent of rainforest and wet and dry land stretching east of the Andes from Colombia to northern Argentina and by a substantial area of water deficient territory on the west south and northeast of the continent Some highly fertile flat subtropical lands about the Parana and La Plata valleys are of minor extent by comparison to the whole South America s resources must be classed largely as potential although the rapidly growing population may bring changes within the twentieth century Like Africa South America has some important mineral deposits chiefly the metals and petroleum

North America Large sections of North American lands benefit from the advantages which characterize middle latitude humid land resources under present technology Disadvantages of desert and semiarid environments on this continent are tempered somewhat by the interspersal of mountain ranges throughout these drier regions Taiga and other northern climatic environments are coincident with the igneous rock in the Laurentian shield and tropical environments are of small extent In sum this continent may be considered to have one of the best balanced sets of resources considering its substantial endowment in minerals of many different kinds extensive forest lands and great and varied agricultural lands and the productive fisheries off both Atlantic and Pacific shores North America has the highest ratio of employed resources to land area of all continents In addition it still contains significant potential resources

Summary comment The earth s resource pattern has certain general characteristics (1) Minerals usable under present technology are found in every environment although mineral types differ according to location in sedimentary or igneous and metamorphic rock areas Mineral exploration will continue indefinitely in all land areas but the mineral resource possibilities of North America and the European part of the Eurasian continent have been examined in greater detail than those of any other large area Ocean basins are the least known part of the world for their mineral possibilities (2) Agricultural lands and forest lands usable under present technology are dominated by those lying in middle latitudes Currently sections of the taiga are becoming more important as forest resources (3) The great potential agricultural and forest resources, if some technological improvement is assumed lie within the tropical environments possibly in northern South America

See CLIMATOLOGY LAKE MINERAL FUEL AREAS MINERAL RESOURCE AREAS RIVER SOIL ZONAL DISTRIBUTION VEGETATION ZONES (WORLD)

[F A A]

Bibliography M R Huberty *Natural Resources* 1960

Earth sciences

Sciences primarily concerned with the atmosphere the oceans and the solid earth They deal with the history chemical composition physical characteristics and dynamic behavior of solid earth fluid streams and oceans and gaseous atmosphere Be-

cause of the three-phase nature of the earth system earth scientists generally have to consider the interaction of all three phases—solid liquid and gaseous—in most problems that they investigate

The geosciences (geology geochemistry geophysics) are concerned with the solid part of the earth system Geology is largely a study of the nature of earth materials and processes and how these have interacted through time to leave a record of past events in existing earth features and materials Hence geologists study minerals rocks ore deposits mineral fuels and fossils and the long term effects of terrestrial and oceanic waters and of the atmosphere They also investigate present processes in order to explain past events

Geochemistry involves the composition of the earth system and the way that matter has interacted in the system through time For example by studying the behavior of radioactive substances it is possible to determine how old the substances are and how much energy has been released through time as a result of decay of radioactive compounds

Geophysics deals with the physical characteristics and dynamic behavior of the earth system and thus concerns itself with a great diversity of complex problems involving natural phenomena For example earthquakes vulcanism and mountain building throw light on the structure and constitution of the earth s interior and lead to consideration of the earth as a great heat engine Study of the magnetic field involves considering the earth as a self sustaining dynamo

The atmospheric sciences commonly grouped together as meteorology are concerned with all chemical physical and biological aspects of the earth s atmosphere Although the study of weather used to be the chief occupation of meteorologists man s entry into the space age calls for a vast increase in knowledge of the environment through which vehicles and ultimately living things will go and return Consequently many aspects of the earth s atmosphere are now being studied intensively for the first time As an example great planetary currents in the atmosphere and also in the oceans are now being investigated not only for the light they may shed on a better understanding of weather but also as a basis for understanding more fully the motion of the entire atmosphere and oceans

Oceanography encompasses the study of all aspects of the oceans—their history composition physical behavior and life content Before World War II little was known about the oceans of the world During that war many important characteristics of the ocean were discovered and since then with instruments and facilities developed during the war oceanographic research has been going on at a quickened pace [R R S]

Earth tides

Cyclic motions a few inches in height similar to and caused by the same forces as daily tides in the sea Tidal motions of the earth's crust are much smaller than ocean tides because the earth and its interior are more rigid but the elasticity does allow a tidal deformation The tidal force which results from the attraction of moon or sun has an effect on gravity so earth tides may be measured by the use of such instruments as gravimeters See GEODESY TERRESTRIAL GRAVITATION

Observations of earth tides Early efforts to measure earth tides date from the end of the nineteenth century accounts are found in George Darwin's *Scientific Papers* and in other writings on natural philosophy of about 1890 Later workers recorded with the aid of an interferometer the changes in water level at the ends of two mutually perpendicular pipes 500 ft long buried in the ground

Later A Marussi also employed instruments of large dimensions for sampling the tilting of a large block of ground He installed in the cave Grotta Gigante near Trieste Italy, two horizontal pendulums of vertical dimension 75 m having a free period of 8 min In Fig 1 is shown a tilt record obtained by Marussi The semidiurnal period with double amplitude about 0 '025 of arc is prominent

(sensi

of ob

years

gravimeters for recording diurnal [illegible] one in gravity to a precision of less than 10 [illegible] times that of the base value of gravity have become available Figure 2 is a reproduction of a gravity record taken in the Belgian Congo during March and April of 1958 under unusually simple lunar solar relationships Because the moon was new on March 20 the prominent lunar and solar semidiurnal tides were then in phase Furthermore both the sun and moon were within a few degrees of the equatorial plane and the moon was at apogee so that its angular velocity in orbit about the earth was constant The amplitude of the semidiurnal gravity tide is large at new moon on March 20 diminishes to a minimum at first quarter on March 28 when the solar and lunar tides are in quadrature and again achieves maximum amplitude at full moon on April 4 3 75 h This exceeds the previous maximum because the moon was then almost at perigee The record shows a maximum double amplitude of 0 000325 gal which is 29% greater than the amplitude for March 20 (one gal is equal to an acceleration of 1 cm/sec^2) Theoretically the increase should be 26% The observed amplitudes are greater by 16 5% than those which would be observed on a rigid earth The significance of this increase will be discussed later

In addition to the two chief instruments the horizontal pendulum and the gravimeter linear strain seismometers are used to measure changes in distance between reference points separated horizontally by 20 m or more These measurements provide independent information about the tidal deformation of the earth Astronomical observations relative to changes in the length of day and concerning the 14-month Chandlerian period of the

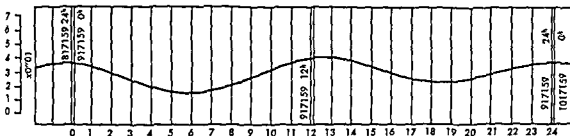


Fig 1 Graph of north south component of tidal tilt at Grotta Gigante Trieste Italy (By A Marussi)

wobble of the pole supply evidence which is closely related in its implications to the evidence provided by observations of tidal tilts and of gravity tides

Potential and gravitational forces Consideration of only the lunar tides will generally suffice for present purposes The semidiurnal solar tidal potential S_2 is about 46 6% of the corresponding lunar potential M_2 It is known that the tide producing gravitational potential of the moon is expressible as the sum of solid spherical harmonics of degree 2 and higher Thus

$$U = U_2 + U_3 + \cdots = \frac{fM}{R}\left[\left(\frac{r}{R}\right)^2 P_2(\cos\theta) + \left(\frac{r}{R}\right)^3 P_3(\cos\theta) + \cdots\right] \quad (1)$$

where f is Newton's constant $6\,67 \times 10^{-8}$ cm^3/(g)(sec)^2 r is the distance of the observation point P from the earth s center O R the distance between O and the moon s center M the mass of the moon θ its geocentric zenith angle MOP (disregarding the slight difference between the true vertical and the geocentric direction OP) and the symbols $P_n(\cos\theta) = P_n(\theta)$ $P_n^m(\cos\theta)$ denote respectively the Legendre polynomials and the associated Legendre spherical harmonics of the first kind A similar equation obviously applies in the case of the sun

Thus for a point at geocentric latitude ϕ on the surface of a perfectly rigid earth of equatorial radius a the formulas for the vertical (upward) component of tidal gravity $\partial U/\partial r$ and for the horizontal component $1/r(\partial U/\partial\theta)$ in the azimuth of the satellite are

$$g_r = \frac{\partial U}{\partial r} = fMC\alpha^3 a^{-2}[2P_2(\theta) + 3\alpha CP_3(\theta) + 4\alpha^2C^2P_4(\theta) + \cdots] \quad (2)$$

$$g_\theta = \frac{-\partial U}{a\partial\theta} = fMC\alpha^3 a^{-2}[P_2^1(\theta) + \alpha CP_3^1(\theta) + \alpha^2C^2P_4^1(\theta) + \cdots] \quad (3)$$

Here $C = 1 - \epsilon \sin^2\phi$ where ϵ is the ellipticity

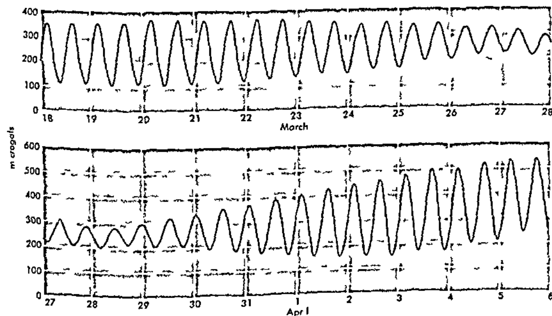


Fig 2 Graphic plot of observed gravity fluctuations at Lwiro Belgian Congo during parts of March and April 1958 (By Edgar Kraut)

so $r = C\alpha$ α is the moon's horizontal parallax aR^{1} For the moon $\alpha^2 < 3\,5 \times 10^{4}$ hence the terms $\alpha^2 P_4$ and higher are generally insignificant For the sun $\alpha_s < 4\,4 \times 10^{5}$ and the first order term $\alpha_s P_3$ is usually omitted

Because the total range of $2P_2$ and of P_2^1 is the same to first order terms the peak to trough amplitudes of the vertical and horizontal components of tidal gravity on a rigid earth would be equal On the real earth the vertical tidal component of gravity is increased by the effects of yielding whereas the horizontal component is diminished

Harmonic constituents In the preceding formulas the two time variable quantities are the horizontal parallax α and the zenith angle θ The parallax varies slowly as the distance to the satellite changes during its orbit The zenith angle has a large diurnal variation as the earth rotates under the satellite The motions and the associated tidal variations are complex and the resultant tidal potential is customarily presented as a sum of many harmonic terms differing in period amplitude and phase In A T Doodson s (1922) tidal analysis 386 harmonic constituents are listed Of these the small group of 10 given in Table 1 is a sample showing the major amplitudes available throughout the large range of periods from 0 34 day to 18 6 years

Table 1 Some harmonic constituents of earth tides

Symbol for constituent	Period lays	Amplitude
M_3	0 3150	0 0119
M_2	0 5175	0 9081
S_2	0 5000	0 4229
K_1	0 9973	0 5305
O_1	1 0758	0 3769
Mf	13 661	0 1564
M_m	27 554	0 0825
Ssa	182 122	0 0729
Sa	365 25	0 0118
Unnamed	6 799 64 (= 18 62 years)	0 0655

Results of harmonic analyses of earth tide records and also of the useful power spectra analyses which large digital computors make feasible will be presented later For information concerning the large subject of tidal analysis the extensive literature should be consulted

Love numbers and examples The Love numbers h k l are defined as follows in terms of the tidal potential U_2 the value g of gravity on the [illegible]

[illegible] are due to the altered [illegible]

the total potential $V_t(a)$ at the surface is represented by the sum of four terms the undisturbed potential of the earth V_0 the decrease in this potential $(\partial V_0/\partial r)\Delta r$ due to the displacement Δr, the direct contribution U_2 due to the satellite and the additional potential kU_2 due to the altered mass distribution On a fluid earth the equilibrium displacement Δr is determined by the requirement that the deformed surface remain equipotential at its original value V_0 Thus $(\partial V_0/\partial r)\Delta r + U_2 + kU_2 = 0$ With $\Delta r = (h/g)U_2$ and $\partial V_0/\partial r = -g$

$$h_f = 1 + k_f \qquad (4)$$

On a homogeneous incompressible earth the change in mass distribution is completely specified by the superficial displacement Δr Setting $\Delta r = \beta r_0 P_2(\cos\theta)$, where β is a small dimensionless constant the potential V at external points caused by the displacement Δr in a sphere of density ρ_e is readily computed to be

$$V(r\,\theta) = 0\,8\pi f\rho_e \beta r_0^5 r^{-3} P_2(\cos\theta) \quad r \geq r_0 \qquad (5)$$

In terms of the Love number k this may be written

$$V(r,\theta) = kU_2(r_0) r_0^3 r^{-3} \qquad (6)$$

an expression which is also valid for a nonhomogeneous symmetrical earth In Eq (5) substitute $\beta r_0 P_2(\cos\theta) = \Delta r = (h/g)U_2(r_0)$ and equate the right hand members of Eqs (5) and (6) obtaining for a nonrotating homogeneous incompressible sphere (for which $g = 4f\rho_e r_0/3\pi$) the relation $k/h = 3/5$ This is valid whether the sphere be solid or fluid For the fluid case Eq (4) with $k/h = 3/5$ yields $h = 5/2$ $k = 3/2$ For the rotating earth the corresponding relation is

$$k/h = 0\,8\pi f r_0 \rho_e g^{-1} \qquad (7)$$

Lord Kelvin's solution for a homogeneous sphere of rigidity μ yields

$$h = \frac{5/2}{1+q} \quad k = \frac{3/2}{1+q} \quad l = \frac{3/4}{1+q} \quad q = 19/2(\rho r_0 g)$$

This study led to his famous assertion that the earth is more rigid than steel

For the case of a rigid incompressible sphere of mean density $\bar{\rho}$ covered by a thin layer of liquid of density ρ Eqs (5) and (7) lead to the relation

$$k/h = 3\rho/5\bar{\rho} \qquad (8)$$

Thus for a rigid earth of mean density 5 517 covered with a shallow ocean of density 1 025 the values for h and k are $h = 1\,125$ $k = 0\,125$ If $\rho = 0$, then $k = 0$ $h = 1\,0$ and the associated tide of height U_2/g is called the equilibrium tide

Observable gravity, tilt, strain The observable total component of gravity on a yielding earth as modified by the tidal potential U_2 is

$$-g_r = \frac{\partial V}{\partial r} = \frac{\partial V_0}{\partial r} + \frac{\partial U_2}{\partial r} + k\frac{\partial}{\partial r}[U_2(r_0) r_0^3 r^{-3}] + \frac{\partial^2 V_0}{\partial r^2}\Delta r \qquad (9)$$

Here V_0 the gravitational potential of the earth satisfies the Laplace equation in its symmetrical form namely

$$\frac{\partial^2 V_0}{\partial r^2} = \frac{-2}{r}\frac{\partial V_0}{\partial r} = \frac{2g}{r}$$

By definition $\Delta r = (h/g)U_2(r_0)$ Hence at the surface $r = r_0$

$$g_r(a) = g_0 - \frac{2U_2(r_0)}{r_0}(1 + h - \tfrac{3}{2}k) \qquad (10)$$

Thus on a yielding earth the amplitude of the gravity tide due to U_2 is larger than that on a rigid earth by the factor $(1 + h - \tfrac{3}{2}k) = \delta$

The level bubble or horizontal pendulum measures a component of tilt of the earth's surface with respect to a level surface The idealized undisturbed surface of the earth is represented by $\bar{r} = a(1 - \epsilon \sin^2 \phi)$ With respect to this datum the surface of the earth is elevated an amount $\Delta r_s = (hU_2/g)$ by the tidal potential U_2 The angle ψ_s in the azimuth of maximum tilt which the disturbed surface makes with respect to its undisturbed counterpart is

$$\tan^{-1}\psi_s = -\frac{\partial(\Delta r_s)}{\partial\theta}d\theta\,[\bar{r}\,d\theta]^{-1} = -r^{-1}\frac{\partial\Delta r_s}{\partial\theta}$$

where θ as before is the zenith angle of the satellite Similarly the level surface through the observing point is tilted by the angle

$$\tan^{-1}\psi_f = -\frac{\partial\Delta r_f}{r\,\partial\theta}$$

so the observable difference angle is

$$\psi_s - \psi_f = -\frac{\partial}{\bar{r}\,\partial\theta}(\Delta r_s - \Delta r_f)$$

when these angles are small But $\Delta r_s = (h/g)U_2(r_0)$ by definition of h, and for the level surface $h = 1 + k$ Thus

$$\psi_s - \psi_f = \frac{1 + k - h}{g}\frac{\partial U_2}{r\,\partial\theta} = (1 + k - h)\frac{g_\theta}{g} \qquad (11)$$

Namely (because $h > k$) the observed tilt angle is smaller than that which would be observed on a rigid earth by the factor $(1 + k - h)$

A strain seismometer measures the relative change in distance $\Delta L/L$ between two points separated at distance L on the earth's surface Taking L and ΔL in the azimuth of the satellite, the change in distance ΔL may be expressed in terms of the Love number l and the respective total surface displacements u_1 and u_2 at the two points as follows

$$\Delta L = u_2 - u_1 = -\frac{1}{a}\frac{\partial u}{\partial\theta}(a\,d\theta)$$

$$= -\frac{1}{a}\frac{\partial}{\partial\theta}\left(\frac{l}{g}\frac{\partial U_2}{\partial\theta}\right)L = \frac{l}{gL}\frac{\partial g_\theta}{\partial\theta} \qquad (12)$$

where g_θ is the theoretical value on a rigid earth of the horizontal component of the gravity tide Thus the strain gage determines l directly in terms of the known quantities g and $dg_\theta/d\theta$

Ratio k/h Recent observations on artificial satellites provide precise values of the leading coefficients in the potential of the earth's external gravitational field The two chief terms in this potential are

$$V = V_0 + V_2 +$$
$$= A_0 r^{-1} - A_2 r^{-3} P_2(\cos\theta) + \qquad (13)$$

where θ is the co-latitude, $A_0 = (3\,98618 \pm 00003) \times 10^{20}\text{cm}^3\text{sec}^{-2}$ $A_2 = (1\,7555 \pm 0001) \times 10^{35}\text{cm}^5\text{sec}^{-2}$ The equatorial bulge produced by the earth's rotation represents a deformation of the same type (but of opposite sign) as that associated with the bodily tides Thus the spheroid is represented to first order terms by $r = r_0(1 - \beta P_2 \cos\theta)$ Corresponding to this deformation are the Love numbers k and h defined above Thus

$$\Delta r = -r_0\beta P_2(\cos\theta) = hg_1^{-1}U_2(r_0\,\theta)$$
$$V_2 = kU_2(r_0,\theta)r_0^3 r^{-3} = -A_2 r^{-3} P_2 \cos\theta$$

Whence $\qquad k/h = A_2 r_0^{-4}\beta^{-1}g_1^{-1} \qquad (14)$

On a nonrotating spheroid, the value of gravity g_1 at the co-latitude for which $P_2(\cos\theta) = 0$ (that is, $\theta = 54\,7°$) is $A_2 r_0^{-2}$
Thus

$$k/h = A_2 A_0^{-1} r_0^{-2}\beta^{-1} \qquad (14a)$$

With the above values of A_2 and A_0, and $\beta = 0022378$ $r_0 = 6\,37114 \times 10^8$ (that is, $\epsilon = 003353$, $a = 6\,37827 \times 10^8$), is obtained

$$k/h = 4848 \qquad (15)$$

for a nonrotating static earth

Values of k and h have been computed for models of static elastic spherical earths by H Takeuchi, by Z Alterman H Jarosch, and C L Pekeris, and by M S Molodenski In the former two studies, the ratio k/h generally is between 458 and 477, whereas for the sixteen models computed by Molodenski this ratio lies between 484 and 507, for ten of these the ratio is 500 ± 005

The equatorial bulge of course represents nearly

ference $C - A$ in the principal moments of inertia of the earth to the mass of the earth, M_e The ratio k/h is thus essentially determined by the ratio $(C - A)\beta^{-1}$ since the value of the denominator $M_e r_0^2$ is relatively well determined Hence, with due

To enable the computation of both k and h from gravity tide observations alone the value for the static case $k_s/h_s = 48$ will be adopted According to Pekeris the dynamic value (12 hr period) for k is $k_D = 1\,080\,k_s$. The dynamic value of h

rotating earth is $h_{DR} = 1\,031(1\,0023)h_s = 1\,0286\, h_s$ Accordingly the static value k_s is

$$k_s = (1\,08)^{-1}k_d = (1\,08)^{-1}(h_{DR}k_D^{-1} - 1\,5)(\delta_D - 1) = 1\,912(\delta_D - 1) \quad (15a)$$

Here δ_D is the dynamic value of δ observed on the rotating earth This formula enables comparison of values k_s deduced from gravity tide observations with the corresponding static values obtained from astronomical observations

Observational results Results of gravity tide observations are generally reported as values of the quantity $\delta = 1 + h - \frac{3}{2}k$ and associated phase lag for the several pertinent tidal constituents For tilt measurements $\gamma = 1 + k - h$ is the quantity reported Together δ and γ determine k and h alternatively knowledge of either δ or γ and of the ratio k h enables determination of k and h At the 1959 IGY earth tide conference in Trieste results of gravity tide observations at 14 stations in western and eastern Europe and in South Africa were reported The minimum duration of observations was 30 days at three observatories records were taken for 248 days or more The reported values of δ for the M_2 tide varied from a minimum of 1 151 to a maximum of 1 238 with a mean of 1 190

Genoa) Their mean value is +1 1° (omitting the +16°4 value) However the significance of the values for phase angles derived from these harmonic analyses is open to question particularly in view of the unfavorable results of test analyses made on assumed tidal functions subject to instrumental drift reported by W Horn and K Rinner

Values for $\delta(M_2)$ and $\delta(M_3)$ and their phases at several additional stations are listed in Table 2 Lwiro and Bunia were among 13 stations occupied for gravity tide observations in a program jointly supported during the International Geophysical Year by the Office of Naval Research and the U S National Science Foundation Of the values for $k_s(M_2)$ probably the most significant are the few in equatorial Africa where the stations were far from the sea Their mean is 0 330 Another determination of k is that of W Markowitz (1959) based on analysis of the lunar tidal terms in the variation of the speed of rotation of the earth He obtains $k = 0\,34 \pm 0\,05$ From the 14 month period of the Chandler wobble, W Munk and G MacDonald obtain $k = 0\,31$, after introducing a correction of $-0\,06$ for the effect of the oceans in the value which H Jeffreys and R Vincente (1957) obtained after correction for dynamic effects of the core The values derived from these independent methods are now in better agreement than was formerly the case The disturbing feature of the earth tide observations is the large values of the phase lags which (as will be indicated in the next section) are much larger than needed to account for the tidal energy dissipation

Tidal dissipation of energy The best estimates of the transfer of angular momentum from the earth to the moon are derived from astronomical data concerning the secular acceleration of the moon in its orbit Studies by W Munk and G MacDonald (1960) based upon observations during the last 250 years of the positions of the moon sun and Mercury lead to the value $2\,7 \times 10^{19}$ ergs/sec for the earth's rate of loss of rotational energy caused by the moon This is nearly twice Jeffreys value (1929) of $1\,4 \times 10^{19}$ which was based upon observations of ancient eclipses Using the larger value the mean lunar retarding torque is

$$T = 2\,7 \times 10^{19}(\widetilde{\Omega} - n)^{-1} = 4\,0 \times 10^{23} \quad (16)$$

where $\widetilde{\Omega}$ is the mean value of the component normal to the line to the moon of the earth's angular velocity and $n = 2\,6617 \times 10^{-6}$ the mean angular velocity of the moon with respect to the earth The former is approximately

$$\tfrac{1}{2}[1 + \cos\,(23°17')]\Omega = 0\,959(7\,2921 \times 10^{-5}) = 6\,993 \times 10^{-5}\ \text{radians/sec}$$

With reference to the loci of tidal energy loss it is known that the atmospheric tides lead the sun and contribute on balance an energy gain at a rate

Table 2 Values for M_3 and M_2 and their phases*

Phase	Bangui 18 56°E 4 4°N L Steinmetz	Lwiro† 28 80°E 2 24°S F Kraut	Bunia† 30 18°E 1 63°N R Forbes	Austin Texas† 262 27°E 30 35°N L LaCoste
	Station coordinates and observer			
$\delta(M_3)$		1 160 ± 015		1 120 ± 015
$k(M_3)$		0 306 ± 03		0 230 ± 03
$\kappa(M_3)$		8 0° ± 4 5		
$\delta(M_2)$	1 173	1 161 ± 0015	1 181	1 123 ± 0015
$k_s(M_2)$	0 331	0 314 ± 009	0 346	0 235 ± 009
$\kappa(M_2)$	1 19°	1 7° ± 3°	−0 90°	
$\epsilon(M_2)$	8 07°	12 0°	−5 9°	

* O[illegible]
† [illegible]
at L[illegible] versity of California
tern [illegible] Austin noise level at

about 8% of the net loss rate 2.7×10^{19} ergs/sec. From the considerations which follow it will appear that the core can contribute only an insignificant fraction of the required loss, that at least a quarter of the loss may occur in the mantle, and that small tides in the deep oceans as yet inadequately measured may easily account for the loss.

The accelerating torque on the moon, or conversely the decelerating torque on the earth due to the potential

$$I_2 = kU_2(r_0, \theta) r_0^3 r^{-3}$$

of the deformed earth is

$$T = MR \frac{\partial I_2}{R\, \partial \theta} = \tfrac{3}{2} k M U_2(r_0, 0) r_0^3 R^{-3} \sin 2\epsilon \quad (17)$$

where ϵ is the angle of lag of the axis of the deformed earth (ϵ is six or eight times greater than the angle commonly used to denote the phase lag of a harmonic constituent of the observed tide). Writing

$$g = g_0(1 + h - \tfrac{3}{2} k) \cos (\omega t - \kappa)$$
$$g - g_0 = A \cos (\omega t - \epsilon)$$

one finds, when ϵ and κ are small,

$$\epsilon/\kappa = \frac{1 + h - \tfrac{3}{2}k}{h - \tfrac{3}{2}k} = 6.8$$

In Eq. (17) introduce for $(r_0 R^{-1})$ and $U_2(r_0)$ their mean values 0.016593 and 3.5132×10^4, respectively, obtaining

$$T = 4.0 \times 10^{-3} = 1.77 \times 10^{-5} k \sin 2\epsilon \quad (18)$$

Because k is approximately 0.29, the required value of ϵ is 2.0°. An estimate of the value of ϵ for the earth may be made as follows. The phase lag P of a linear oscillator of natural frequency f_0 and damping coefficient 2η excited at frequency f is given by

$$\tan P = f_0^2 Q^{-1} (f_0^2 - f^2)^{-1} \qquad Q^{-1} = \frac{2\eta f}{2\pi f_0^2} \quad (19)$$

Here Q^{-1} has its usual significance as a dimensionless measure of the loss rate in a vibrating system. The value of Q^{-1} for inorganic nonferromagnetic solids is nearly independent of frequency in the range 1×10^{-7} to 1×10 cps (Knopf and MacDonald, 1958). Studies of long-period earthquake waves and laboratory studies of rocks indicate that the value of Q^{-1} for rocks is approximately 100–500 for seismic waves having periods in the range 1–4 min. These values are consistent with the values of Q^{-1} for rock samples measured at frequencies of 140–4500 cps in the laboratory. Because the natural period of the earth in the P_2 mode is about 0.9 hour, the ratio $(f/f_0)^2$ in the case of the semidiurnal tide is very small. With Q^{-1} constant at 100, Eq. (19) gives for the phase lag P the value 0.01 radians = 0.573°. Thus tides in the solid earth might reasonably be expected to dissipate the fraction $(0.573)(2.0)^{-1} = 0.29$ of the required lunar tidal power. With a Q^{-1} value of 25 at the 12-hour tidal period, the crust could dissipate all the lunar tidal energy.

To examine the possibility of dissipation in the core, it is convenient to write Eq. (17) in the alternative form

$$T = 1.2\pi \beta r_0^5 \rho / MR^{-3} \sin 2\epsilon \quad (20)$$

At the earth's surface, the discontinuity in the (effective) density is 4.598, which is about the same as that at the core boundary. Thus for a given value of β the tidal torque available from a core of radius $0.55a$ is reduced by the factor $(0.55)^5 = 0.05$. Hence it is difficult to see how the core can contribute more than a small fraction of the loss attributable to the mantle.

Consider next the order of magnitude of the energy dissipation in a shallow uniform sea covering a rigid earth of mean density 5.516. In the preceding section on Love numbers, it was noted that the values for h and k in this case are $h = 1.125$, $k = 0.125$. With these values the tide height is 40.4 cm, and a phase lag of 5.2° or about 21 min is required to account for the entire loss rate 2.7×10^{19} ergs/sec. At the most favorable time lag of 3 hours the height of the tidal bulge required to account for the dissipation is only 7.3 cm. Thus tides of long wavelength and modest amplitude with phase lags of 1–3 hours could account for most of the tidal energy loss, aided possibly by significant contributions from the mantle. G. Groves and W. Munk have made a calculation based on G. Dietrich's data on tide heights of the work done on the oceans by the moon and sun and arrive at a total of 4.2×10^{19} ergs/sec, which somewhat exceeds the required total of 3.2×10^{19} ergs/sec.

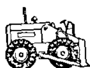
It is clear that deformations expressed by solid harmonics of degree higher than two can contribute essentially nothing to the tidal torque because these terms contain an additional factor $a/R = 0.0166$. To account for the tidal rate of energy loss of approximately 2.7×10^{19} ergs/sec, it is necessary to postulate the existence of second-degree solid harmonics of suitable amplitude and phase lags in either the oceans or the mantle, or both. To this extent the gross geometrical features of the required deformations are clear, but the detailed mechanisms and loci of energy dissipation remain to be investigated. [L.B.SL.]

Bibliography. S. Fluegge (ed.), *Handbuch der Physik*, vol. 48, 1957; H. Jeffreys, *The Earth*, 4th ed., 1959; W. H. Munk and G. J. F. MacDonald, *The Rotation of the Earth*, 1960; P. Schureman, *Manual of Harmonic Analyses and Prediction of Tides*, USCGS Spec. Publ. 98, 1941.

Earthmover

Any of a variety of construction machines designed to move or transport earth. Earthmovers include heavy-duty trucks with high-sided dump bodies, self-propelled or towed scrapers, wagons, and bulldozers as illustrated. The bulldozer mounted on a wheeled or crawler tractor is suitable for moving large quantities of earth for distances of several hundred feet. Scrapers and wagons are efficient for

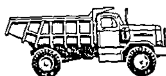
moving earth over relatively level terrain for distances up to one or two miles For longer distances or for grades in excess of 5% trucks are the most practical Scrapers have the advantages of being able to load themselves without help from a crane



(a)



(b)



(c)

Typical earthmovers (a) Tractor dozer (b) Bottom dump trailer (c) Rear dump truck (Cummins Engine Co)

or power shovel and of discharging their loads in finely controlled layers Trucks can be of either the side or rear dump type Wagons can have side bottom or rear dump mechanisms *See* BULK HANDLING MACHINES CONSTRUCTION EQUIPMENT

[F M Y]

Earthquake

A series of elastic waves propagating in the earth These waves are to be distinguished from those set up by vibrating machinery and from those continuous vibrations called microseisms The nature of the transient disturbance at the source of an earthquake is debatable The American view also held by many Europeans is that the source of the vast majority of shocks is the movement of the earth along a plane of weakness called a fault (*see* FAULT AND FAULT STRUCTURES) The rubbing together of the two fault surfaces generates the elastic waves The reason for the fault movement is presumed to be a slow accumulation of strain in the rocks over a considerable region until they fail and rupture (*see* ROCK MECHANICS) Thus energy slowly accumulated is suddenly released a compensation considered under the elastic rebound theory (*see* SEISMOLOGY) As to the source of the gradual accumulation of strain there is little agreement A number of possibilities have been suggested including shrinking of the earth isostatic readjustment plastic currents in the mantle below the earth s crust and radioactive heating

American seismological thinking has been strongly influenced by the 1906 California earthquake which accompanied a break in the earth s ... of on the San Andreas Fault some 270 miles ...

in particular other types of source a ... monly preferred by many such as any sudden changes in size and shape of material at depth It is true that the great majority of earthquakes are not accompanied by conspicuous surface faulting

Classification Efforts to classify earthquakes have not been satisfactory One type is called tectonic This means having to do with forces in the earth s crust which form mountains (*see* OROGENY) Earthquakes caused by fault breaks are called tectonic earthquakes Other classifications have used the words volcanic and plutonic A shock was called volcanic if associated closely with volcanic activity in time and space Some scientific writers however, have said that if such a shock is felt over a very large area it must be tectonic Again some do not care to use the word tectonic to describe earthquakes originating at great depths—below the earth's crust—and instead have used plutonic Insufficient knowledge of earthquake sources hinders development of a genetic classification

Analysis of the first displacements in dilatational (*P*) and shear (*S*) waves as recorded on seismographs leads to knowledge of the forces acting at the source Such studies have indicated that a single couple (with moment) or a double couple in the same plane (one set of forces at right angles to the other) explains the nature of the beginnings of (*P*) waves There is some indication that a double couple better explains the beginning of (*S*) waves *See* SEISMOGRAPH

Effects When the elastic waves set up by a great earthquake arrive at the earth's surface they produce varied effects Landslides are common in hilly regions particularly if the shock was preceded by a rainy season River bottom land lurches about leaving wide cracks Man made filled ground suffers greatly by being thrown about The flow of springs is altered usually only temporarily Aquifers are frequently compressed causing temporary fountains, sometimes compression of underground strata causes eruption of sand and gas-sand blowouts Poorly built structures on filled land suffer extremely, whereas buildings on rock survive much better

Large earthquakes centering under the ocean or beneath coastal lands set the water into three types of motion First movement of the ocean bottom or submarine landslides caused by shaking displaces the ocean's surface This displacement sets up gravity waves in the water which travel at speeds as great as 475 mph depending on the depth of the water These have wave length measured in tens of miles and pass unnoticed by observers on ships at sea As they approach coasts however, their heights increase They frequently cause great damage in harbors particularly in



Fig 1 Faulting developed during Fairview Peak earthquake of December 16 1954

shallowing bays with broad openings toward the sea Under such conditions waves as great as 90 ft in height have been reported These waves flood the land and destroy lives and property They are popularly known as tidal waves but are technically called tsunamis (*see* TSUNAMI) Tsunamis fre quently set bays into their own natural periods of water oscillation which may continue for days These oscillations are among the water movements called seiches (*see* SEICHE) They may also be set up by severe shaking as in Loch Lomond at the time of the Lisbon earthquake In the second type of motion the first preliminary waves (waves of compression rarefaction) propagate in the water killing fish and sometimes wrecking ships Third such waves may become entrapped in a low velocity layer in the ocean and propagate for vast dis tances Such a layer exists in certain latitudes at some seasons of the year Such waves called *T* waves by seismologists record late as high fre quency waves on seismograms at stations not far from coasts

Foreshocks and aftershocks A large earth quake is occasionally preceded by small quakes (foreshocks) during a few hours or days whereas such an earthquake is always followed by countless aftershocks These aftershocks are unnerving and are often large enough to cause additional dam age An aftershock sequence may last a year Rarely do two or three large earthquakes of equal size occur in the same area within a short time

Size of earthquakes Two measurements of earthquakes are in common use intensity and magnitude

Intensity This attribute of an earthquake is de fined by its effects on the earth's surface by such results as landslides and cracks in ground and through its effects on man and man made struc tures Thus the intensity of a great submarine earthquake may be said to be zero On the other hand a much smaller earthquake centering near a town poorly built on filled ground will have a very [illegible]

struction of buildings and the distance from the source If man chooses the location of his cities care fully and builds excellent structures the intensity of earthquakes can be greatly reduced In the lower ranges of intensity the criteria for rating are its effects on people in such ways as general alarm inducing fright and awakening those asleep

Magnitude Such measure of an earthquake is based on the ground motion as recorded by seismographs It is expressed in amounts related to the energy given out at the source The relationships used express the logarithm of the energy as equal to a constant plus another constant times the magni tude Several values of the constants have been sug gested The increase of a unit in magnitude thus indicates an increase in energy of order 10 A given earthquake has only one magnitude whereas its intensity in various localities varies with the factors cited above

Historical earthquakes To be a great historical earthquake the shock needs of course to have been in a region where some records were kept It is of

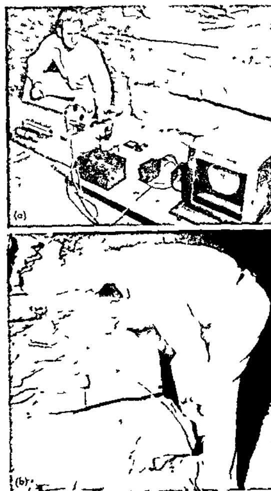
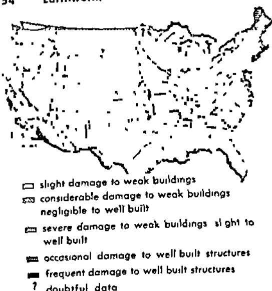


Fig 2 Field installation of a seismograph at Mammoth Cave Kentucky (a) Recorder (*left*), amplifying unit (*right*) (b) Pickup unit which detects the earth vibrations is located a short distance from the units (*U S Coast and Geodetic Survey*)



▭ slight damage to weak buildings
▭ considerable damage to weak buildings negligible to well built
▭ severe damage to weak buildings slight to well built
▬ occasional damage to well built structures
▬ frequent damage to well built structures
? doubtful data

Fig 3 Sketch map showing probabilities of quake risk in the United States (*After C F Richter News paper Enterprise Association Services Inc*)

more interest if centered in a region which was or is now heavily populated

The most famous earthquake probably was the Lisbon earthquake of November 1 1755 Three great shocks rocked Lisbon during the morning The death toll in Lisbon was between 30 000 and 70 000 people All large public buildings and most of the churches were ruined The main shock which occurred at noon lasted 6 or 7 minutes an unusually long time The shock was strong throughout Portugal and Spain and did great damage at Fez and Mequinez in Morocco The earthquake was observably felt over about 1 500 000 mi²

The New Madrid earthquakes which centered in southern Missouri in 1811 and 1812 were the most severe on record within the United States Three of the series occurring on December 16 January 23 and February 7 were very great In a region of some 50 000 mi the ground was greatly disturbed the soil was fissured and thrown about domes were uplifted and forest trees were broken off or uprooted An area called the Sunken Country, some 40 miles wide and 140 miles long sank 3–9 ft Chimneys were thrown down in Cincinnati 400 miles away The shock was felt along the Atlantic seaboard along the Gulf Coast at the headwaters of the Arkansas and Missouri rivers and in Canada

The California earthquake of April 18 1906 is famous primarily for its great surface fault break (270 miles long) Damage was great near the fault in a number of cities San Francisco San Jose Salinas and Santa Rosa The great fire which followed the shock in San Francisco completed the ruin The death toll was about 1000 Property loss caused by fire and the earthquake has been estimated at $325 000 000 The relationship of geologic foundation to damage was most conspicuous in San Francisco There was little earthquake damage on the rocky hills, more in the valleys between the spurs, much on sandy areas, and very great damage on filled ground This shock was felt north to Coos Bay Oregon, south to Los Angeles and east to Winnemucca, Nevada—not a large area

Seismicity. Certain portions of the earth's crust are more subject to earthquakes than others Wherever high mountains border deep seas there are earthquakes The great encircling zone which surrounds the Pacific Ocean is the source of most shocks Another seismic zone, roughly a great circle includes the East Indies, the Himalayas the Caucasus the Alps and the West Indies [P BY]

Bibliography J Gilluly, A C Waters and A O Woodford *Principles of Geology*, 1957, C F Richter, *Elementary Seismology*, 1958

Earthworm

A common name applied to several terrestrial forms of the class Oligochaeta, phylum Annelida Most of them are world wide in distribution In the United States there are 3 families of earthworms and about 12 species Most American earthworms are less than 12 in in length However there is an Australian species which grows from 9 to 11 ft long and an Ecuador species which is almost as large

Importance. Besides their utilization as fish bait earthworms are important in aereating the soil and in providing passages for the entrance of ground water into the soil They also enrich the topsoil by bringing up parent materials from the subsoil and by carrying humus down to the lower levels Because of their relationship to good soil earthworms have been called the most important animals in the world

Earthworm farming is a business enterprise that is now thriving in several localities Worms are cultured in great quantities for fish bait and are also used by farmers to restock the soil where worm populations have been depleted Several species are reared because the species and the soil type must be matched to maintain strong populations

The earthworm is generally studied in the laboratory in zoology classes It is readily available and is an excellent example of the organization of a triploblastic segmented invertebrate The best known American and European species is *Lumbricus terrestris* and it is this animal that is usually described in zoology texts

Structure. *Lumbricus terrestris*, often called the night crawler, is the largest American species reaching a length of 12 in It is long cylindrical and pointed at both ends The body is marked off into about 180 segments, visible externally as pronounced grooves and internally marked by septa which divide the body cavity into semi independent compartments The terminal mouth and anus are connected by a straight but somewhat modified gut Back of the mouth is a stout muscular pharynx followed by a narrow esophagus leading into a

thin walled enlarged crop The latter leads into the highly muscular gizzard which in turn opens into the intestine unmodified except for a mid dorsal fold the typhlosole that serves to increase the absorptive surface of the intestine This system is admirably suited to extract the maximum nourishment from the quantity of dead vegetation and organic debris which the earthworm eats

The nervous system consists of a small bilobed brain connected to the anteriormost of the ventral nerve ganglia by a small pair of nerves the circumpharyngeal connectives The ventral nerve cord continues with segmental ganglia the length of the worm The circulatory system is of the closed type with five hearts or loops connecting the dorsal and ventral blood vessels There are additional vessels running the length of the worm with segmental vessels branching at right angles to them in each segment The excretory system is well developed each segment except the first three and the last one having a pair of nephridia Each nephridium consists of a ciliated funnel opening into the coelom a coiled tube penetrating the septum leading into the next posterior segment and opening laterally by a small *nephridiopore* near the ventral row of setae

Locomotion Movement is accomplished by setae and by muscular contractions the body wall having both longitudinal and circular muscles The setae are short stiff bristles arranged in series with one pair per segment on each lateral body wall and another pair located ventrolaterally on either side of each segment The body wall outside the muscles is completed by a thin epidermis covered by a transparent cuticle

Reproduction A prominent external feature is the clitellum a glandular enlargement forming a nearly complete band on somites 31 through 37 which plays an important role in the reproductive process Each worm is hermaphroditic and cross fertilization is practiced In mating the ventral surfaces of the pair are applied with the heads in opposite directions This permits sperm from the seminal vesicle of each worm to be transferred to the seminal receptacles of the other In sperm exchange they are held together by a pair of mucous bands secreted by the clitellum of each worm After separating following mating each worm secretes a mucous band around itself and crawls back out of this band depositing in it a few eggs and the recently received sperm The slime band drops off the worm as a lemon shaped egg cocoon about 7 mm long This cocoon is left on the soil where it falls In a few days a single young worm escapes

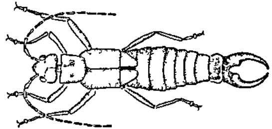


The earthworm *Lumbricus terrestris* length up to 12 in (From T I Storer and R L Usinger *General Zoology* 3d ed McGraw-Hill 1957)

the other eggs acting only as nurse cells In related species several young may develop from each cocoon Several cocoons are deposited each year *See* OLIGOCHAETA [J D B]

Earwig

Any member of the insect order Dermaptera Earwigs are readily recognized by the prominent pair of pincerlike structures on the end of the abdomen Earwigs may grow to 1½ in in length They are



elongate with a prominent abdomen and may be wingless or bear two pairs of wings Usually the forewings are short and leathery covering only the front part of the abdomen the hindwings are membranous They have chewing mouthparts There are about 1100 species but only a few occur in the United States They are of little economic importance except for the European earwig *Forficula auricularia* an introduced species which is widespread on the Atlantic and Pacific Coasts and has become a garden pest eating plant material Earwigs frequently invade houses where they hide during the day in pantries closets or among dishes

The earwig received its common name from the European superstition that it will crawl into the ear of a sleeping person *See* DERMAPTERA

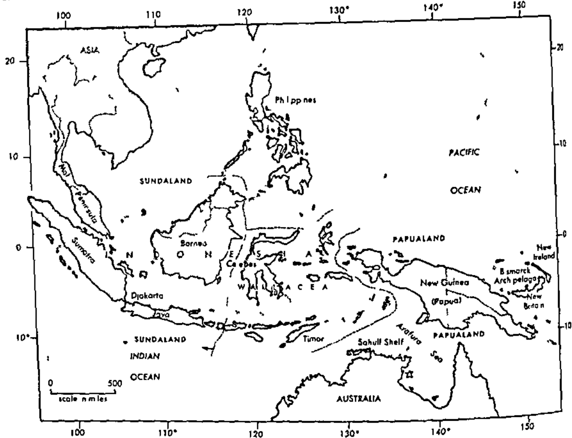
[J D B]

East Indies

The Indonesian Islands proper and for convenience of presentation New Guinea and the Bismarck

The islands are the peaks of mountains formed by folds in the edges of the continental shelves of the two continents Three basic divisions are recognized (1) Sundaland or the islands west of the Macassar and Lombok Straits (2) Papualand or the islands east of the Aroe or Aru Islands and New Guinea and (3) Wallacea or the islands be-

mountains of Sumatra and Java represent folds against the edge of the Sunda Platform The core



East Indies map of locations

area of Papualand is the Sahul Shelf which is an extension of the Australian continent into central New Guinea The volcanic mountains and folds of central and northern New Guinea represent recent folds against this shelf The Sunda Platform and the Sahul Shelf are relatively free from earthquakes and have not had volcanic activity since ancient geological times Wallacea is highly unstable with active volcanoes and frequent earthquakes Some of the islands show evidence of considerable recent uplift The orogenic processes which have produced these folds have been active since the beginning of the Cretaceous *See* CRETACEOUS OROGENY

During the Pleistocene a lowering of the general level of the world s oceans resulted in the emergence of dry land between the western edge of Wallacea and Asia and on the east exposed portions of the Sahul Shelf beneath the Arafura Sea The islands of Wallacea however remained separated by deep water which was an effective barrier to the migration of plants and animals Thus the flora and fauna of Sundaland are closely related to the flora and fauna of Asia while the flora and fauna of Papualand show a relationship to that of Australia and a separation from Asia since early Cretaceous Because of this isolation and because of the great variation of topography soils rainfall humidity and altitude within short distances favorable conditions have existed for the development of numerous species of plants

Since almost all of the islands lie in a belt within 10° of the Equator an equatorial climate prevails with high temperatures throughout the year except at higher elevations The diurnal variations are greater than the range in mean temperatures of the hottest and coldest months Monsoons representing an interplay of air masses between Asia and the Southern Hemisphere control the seasonal change in winds and the variation in precipitation A tropical rainy climate with no less than 2 4 in of rain each month prevails in Sumatra western Java Borneo Celebes and Ceram but a winter dry season occurs in eastern Java and the Lesser Sundas High relative humidity is normal in the low lands

Borneo Borneo or Kalimantan with an area of 282 000 square miles is the third largest island of the world It is a stable portion of the Sunda Platform whose shallow seas contrast with the 12 000-ft depth of the Celebes Sea on its northeast side and the deep Macassar Strait on its east Mt. Kinabalu 13 445 ft in British North Borneo is the highest mountain on the island Low mountain ranges with few peaks over 6000 ft and crossed by easy passes extend southward from British Borneo Nearly half of the island is made up of lowlands under 600 ft in elevation which are frequently swampy Ancient crystalline rocks part of the Sunda Platform are exposed in the mountains but about two thirds of the island is covered with Ter

tiary stratified deposits which contain coal and petroleum These are sometimes overlaid with Quaternary sediments Major oil pools are found near Balikpapan Banjermasin Tarakan and Brunei Diamond gold and silver deposits are known

Crossed by the Equator, Borneo has a tropical climate with little range in temperature coastal stations varying between 77 and 79°F average monthly temperatures throughout the year Annual rainfall in the lowlands varies from 90 to 130 in This is evenly distributed throughout the year without a distinct wet or dry season except in the north and southeast where there is a slight tendency toward a short dry season

A tropical evergreen forest covers large areas of the island but there are also extensive areas of savanna grass and scrub trees which are a result of past agriculture and man made fires

Celebes Celebes or Sulawesi is an irregular or K shaped island with a 3000 mile coast which encloses an area of approximately 72 000 square miles The island is rugged with little land under 600 ft in elevation in fact most of the land is over 1500 ft a peak of the Quarles Mountains rises to an elevation of 11 296 ft Celebes represents one of the least stable portions of the earth s surface Peneplained in the Oligocene Celebes experienced frequent vertical movements in the Late Tertiary and in the Quaternary Post Pleistocene coral reefs are now located at elevations up to 3000 ft Numerous graben lakes occupy depressions along the major fault lines One of these Lake Poso is 5000 ft deep Although there are no important mining activities indications of nickel gold iron and petroleum have been found Hot unvarying temperatures heavy rainfall and high relative humidity result in a heavy cover of tropical rainforest

Timor Timor 300 miles long and varying from 10 to 60 miles in width has an area of about 7300 square miles The island is composed of ancient crystalline rocks and formations representing Paleozoic Mesozoic and Tertiary with some Quaternary These formations are much folded and contorted Violent action occurred in the Quaternary as shown by coral reefs of that age found today at elevations of 4000–5000 ft The numerous volcanoes are extinct the last eruption was in the seventeenth century

Java Java or Djawa one of the richest agricultural islands in the tropics is 670 miles long and varies from 24 to 120 miles in width with an area of 48 800 square miles The mountains which are most prominent in the southern portion of the island were formed by a Late Tertiary fold pushing against the Sunda Shelf which now is the foundation of the northern lowlands Intense volcanization at the end of the Pliocene and Pleistocene resulted in the formation of more than 20 volcanoes of which 13 are currently active New eruptions add to the deep black soils and give them a higher fertility than is normal for the climatic conditions Petroleum is found in a belt running from Semarang through the island of Madura

Lying close to the Equator between 6 and 9°S latitude the lowlands have high even temperatures Djakarta s mean monthly temperature ranges from a minimum of 78°F to a maximum of 79 7°F and its average relative humidity is 78% during the driest month Occasional frosts occur at elevations above 4500 ft The heaviest rainfall is in the mountains and the western lowlands Eastern Java has a pronounced dry season from July through September when the southern monsoon loses its moisture on the southern slopes of the mountains

Sumatra Sumatra the large western island of Indonesia is about 1600 miles long and 250 miles across at the widest portion It has an area of 166 800 square miles and extends from 6°N to 6°S latitude The island is made up of three distinct physiographic sections Youthful folded mountains on the west push against the Sunda Platform Peaks of these mountains rise from 3000 to 8000 ft East of the mountains there is a belt of low folded hills of Tertiary rock structures which contain coal and petroleum deposits The eastern half of the island is composed of extensive swampy alluvial lowlands formed on the Sunda Platform Important petroleum deposits are found near Palembang and Medan Small islands to the east of Sumatra—Lingga Bangka and Billiton—contain major tin deposits

Most of Sumatra has an annual average rainfall of over 150 in The western slopes of the mountains may have as much as 30 in more but the intermontane basins receive less The seasons of less rain which correspond with the high sun period of the hemisphere are not severe enough to interrupt growth This results in a tropical rainforest in the uplands and a tropical swamp forest in the lowlands

New Guinea New Guinea or Papua is nearly 1500 miles long and 390 miles wide at its maximum With an area of approximately 340 000 square miles New Guinea is the second largest island in the world Politically it is divided along the 141st meridian West of the meridian the island is known as Netherlands New Guinea or Irian and east of the meridian it is divided into Papua to the south and Northeastern New Guinea to the north

The southern lowland area is part of the Sahul Shelf and an extension of the Australian continent The high mountains of central New Guinea are a geanticline or broad uplift which has folded against the Sahul Shelf (*see* GEOSYNCLINE) Peaks of the central mountains rise to more than 16 000 ft in the Nassau and Orange Ranges Since the snow line is at about 14 500 ft many of these are snow capped although within 5° of the Equator North of the central mountains there is a synclinal valley with a coastal range to its north Currently and potentially important deposits of minerals have been found gold petroleum coal iron ore tin copper platinum manganese osmiridium and others

Heavy rains are received from both the s winter and summer monsoons This with

temperatures (the mean monthly high in December is 78°F and the low in August is 74°F) results in a tropical rainforest over most of the island except where savannas have been culturally induced

New Britain New Britain or Neu Pommern is the largest island of the Bismarck Archipelago It is 370 miles long and 90 miles wide at its maximum width and it has an area of about 13 000 square miles It is a continuation of the youthful folded northern coastal range of New Guinea Some of the mountains of the rugged range are over 7000 ft in elevation and are active volcanoes Rabaul the capital of the Bismarcks is located in the caldera of an active volcano one of whose walls is broken down and open to the sea

New Ireland New Ireland or Neu Mecklinberg is part of the Bismarck Archipelago The island has an area of 3340 square miles it is 200 miles long with an average width of 20 miles It is very mountainous but lacks active volcanoes [C A M A]

Bibliography O W Freeman (ed) *Geography of the Pacific* 1951 P Honig and F Verdoorn (eds) Science and scientists in the Netherlands Indies *Natuurk Tijdschr Ned Indie* vol 102 spe

Ebenales

A relatively small order of the plant subclass Dicotyledoneae including 4 families with 54 genera

advanced characters with the frequent occurrence of indefinite numbers particularly of stamens The sapodilla family (Sapotaceae) much the largest

ver bell (*Halesia*) the snowdrop bush (*Styrax officinalis*) source of storax (a resin) and *S benzoin* which yields aromatic benzoin gum used in soaps lotions and toothpaste *See* CHICLE EBONY *see also* DICOTYLEDONEAE EMBRYOPHYTA PLANT KINGDOM TREE [P D S]

Ebony

A genus *Diospyros* of the ebony family containing more than 250 species Some species are important for their succulent fruits such as date plum kaki plum and persimmon and several for their timber particularly the heartwood which is the true ebony of commerce

Although it is popularly supposed to be a black wood most species have a heartwood that is only streaked and mottled with black The heartwood is very brittle breaks with a conchoidal fracture and is difficult to work but it has long been in demand The sapwood is white becoming bluish or reddish when cut *See* WOOD (ANATOMY AND IDENTIFICATION)

The use of ebony can be traced to the early Egyptians who probably obtained it in Abyssinia Ebony from India was known to the Greeks before 350 B C At present black ebony is used for knife handles piano keys, finger boards of violins hair brush backs inlays and marquetry Some of the woods called ebony in the trade however belong to different families especially the pulse family Leguminosae



Common persimmon *Diospyros virginiana* (A H Graves *Illustrated Guide to Trees and Shrubs* Harper 1956)

Persimmon *Diospyros virginiana* of the southeastern United States is one of numerous tropical or subtropical species Usually a medium sized tree with black chunky bark it is known to attain 100–125 ft in height with a trunk 20–30 in in diameter The sapwood is in demand for the manufacture of weaving shuttles and heads of golf clubs The fruit is sweet and edible when slightly overripe but when immature it is extremely pun

are too small al of king ticles

of turnery and carving *See* FOREST AND FORESTRY TREE [A H C]

Echeneiformes

The remoras or sharksuckers which form a distinctive although small order of actinopterygian fishes They are also known as the Discocephali



Sharksucker *Echeneis naucrates* length to 38 in (After D S Jordan and B W Evermann *The Fishes of North and Middle America* U S Natl Museum Bull 47 1900)

The spinous dorsal fin characteristic of typical perciform fishes is modified into an oval flattened adhesive disk with roughened transverse laminae it lies on the head and nape The terete body has small cycloid scales the pectoral fin is placed high on the side and the pelvic fin is thoracic

This order (or suborder as it is ranked by some authorities) includes a single family Echeneidae 4 Recent genera with 8–9 species and one Oligocene genus They attach themselves at will to sharks marlins or other large fishes as well as to porpoises and sea turtles They travel widely with these commensals The species are distributed in all temperate and tropical seas *See* ACTINOPTERYGII [R M B]

Echinacea

A superorder of Euechinoidea having a rigid test the periproct within the apical system keeled teeth a complete perignathic girdle and branchial slits J Durham and R Melville (1957) include five orders in this group These were formerly distributed among the Stirodonta and Camarodonta in the classification of R Jackson (1912) *See* ARBACIOIDA, ECHINOIDA, EUECHINOIDEA HEMICIDAROIDA PHYMOSOMATOIDA TEMNOPLEUROIDA [H B F]

Echinococcosis

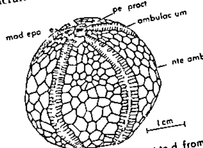
The infestation by the hydatid of *Echinococcus granulosus* also known as hydatidosis The adult worm occurs in the gut of dogs and other canines and the hydatid occurs primarily in the liver and lungs of herbivores and occasionally in man The outer wall of the hydatid composed of stratified layers is produced by the host whereas the inner surface produced by the parasite is a layer of germinal tissue This layer produces brood capsules which discharge into the cavity of the mother cyst The scoleces are produced mainly in the brood capsules The contents of a large cyst feel gritty and are frequently called hydatid sand The hydatid grows slowly but continuously and may attain a volume of 10 or more quarts Should scoleces or brood capsules escape from the mother cyst they can establish new foci of infection

Human infestations are found throughout the world in nontropical areas but are particularly common in certain areas of South America Australia New Zealand and Africa where dogs sheep and cattle are abundant and in intimate association The normal cycle of the tapeworm involves the dog and the herbivore but man may become infected directly from the dog feces or from fecally contaminated food or water Human infestations are rare in North America except in the far north Here the parasite normally occurs in wolves and moose and other wild herbivores but dogs become infested when eating offal from the herbivores Diagnosis of hydatidosis is made by skin tests and x ray examination Treatment is by surgical removal of cysts Iceland which had a heavy incidence of human infection has shown that this parasite can be controlled if proper sanitary principles are observed *See* CYCLOPHYLLIDEA [R S FR.]

Bibliography A C Chandler *An Introduction to Parasitology* 9th ed 1955

Echinocystitoida

An extinct order of Perischoechinoidea which arose in the Ordovician and seems to have inhabited calm shallow lagoons There was no perignathic girdle the number of columns of plates was variable the ambulacral plates overlapped adorally and the interambulacral plates overlapped one another toward the apex and also overlapped the adjacent ambulacrals (see illustration) The shape of the



Aulechinus grayae an echinocystitoid from the Upper Ordovician of Scotland

test varied from spherical (*Echinocystis*) to flattened pentagonal (*Hyattechinus*) The largest forms (*Fournierechinus*) reached a diameter of 30 cm All were extinct by the close of the Permian but before extinction they had given rise to the Archaeocidaridae from which all surviving echinoids probably stem *See* CIDAROIDA PERISCHOECHINOIDEA [H B F]

Echinodermata

A phylum of exclusively marine coelomate animals distinguished from all others by structural peculiarities of the skeleton and coelom The living representatives are the sea urchins (Echinoidea) sea cucumbers (Holothuroidea) sea stars (Asterozoa) and the feather stars and sea lilies (Crinoidea) The distinctive features of living and fossil echinoderms alike are (1) an internal skeleton composed of numerous independent crystalline calcite plates and (2) conversion of a part of the coelom into a water vascular system that is fluid filled vessels which push out the surface of the body as a series of hollow tentacles The water vascular system is an adaptable multipurpose structure which serves the needs of locomotion respiration nutrition or sensory perception All living echinoderms are radially symmetrical as adults but this is not a fundamental character of the phylum for it is not seen in certain fossil forms and is imperfect in others *See* ENTEROCOELA EUCOELOMATA

Echinoderms have existed since the [Cam]brian period During this vast time

500 000 000 years several divergent structural pat terns have evolved so that the surviving groups show few resemblances to the original stock This article refers only to features widely shared among echinoderms For more specific information on any included group use the taxonomy chart to determine the cognate articles Information given in one article will not be repeated in a subsidiary context

Taxonomy of the Echinodermata

Phylum Echinodermata
- Subphylum Pelmatozoa
  - Class Carpoidea
  - Class Edrioasteroidea
  - Class Cystidea
    - Order Diploporita
    - Order Rhombifera
  - Class Blastoidea
  - Class Eocrinoidea
  - Class Paracrinoidea
  - Class Crinoidea
    - Subclass Inadunata
    - Subclass Camerata
    - Subclass Flexibilia
    - Subclass Articulata
- Subphylum Eleutherozoa
  - Class Holothuroidea
    - Order Elasipoda
    - Order Aspidochirota
    - Order Dendrochirota
    - Order Molpadonia
    - Order Apoda
  - Class Asterozoa
    - Subclass Somasteroidea
    - Subclass Asteroidea
      - Order Platyasterida
      - Order Hemizonida
      - Order Phanerozonia
        - Suborder Paxillosa
        - Suborder Notomyata
        - Suborder Valvata
      - Order Spinulosa
      - Order Forcipulata
    - Subclass Ophiuroidea
      - Order Stenurida
      - Order Oegophiurida
      - Order Ophiurida
        - Suborder Euryalae
        - Suborder Ophiurae
  - Class Ophiocystioidea
  - Class Echinoidea
    - Subclass Perischoechinoidea
      - Order Bothriocidaroida
      - Order Echinocystitoida
      - Order Palaechinoida
      - Order Cidaroida
    - Subclass Euechinoidea
      - Superorder Diadematacea
        - Order Diadematoida
        - Order Echinothurioida
        - Order Pygasteroida
      - Superorder Echinacea
        - Order Hemicidaroida
        - Order Phymosomatoida
        - Order Arbacioida
        - Order Temnopleuroida
        - Order Echinoida
        - Order Holectypoida
        - Order Clypeasteroida
        - Order Nucleolitoida
        - Order Cassiduloida
        - Order Holasteroida
        - Order Spatangoida

Morphology. Echinoderms show common ana tomical features with respect to the skin skeleton symmetry alimentary system, reproductive organs, coelom water vascular system and nervous system

Skin The skin is covered by thin ectoderm and often is ciliated The ectoderm disappears during growth in ophiuroids and exposes the skeleton but this is a secondary condition The deeper layers of the integument are mesodermal and comprise mus cular tissue and skeletal tissue

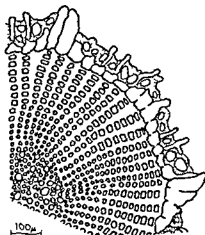
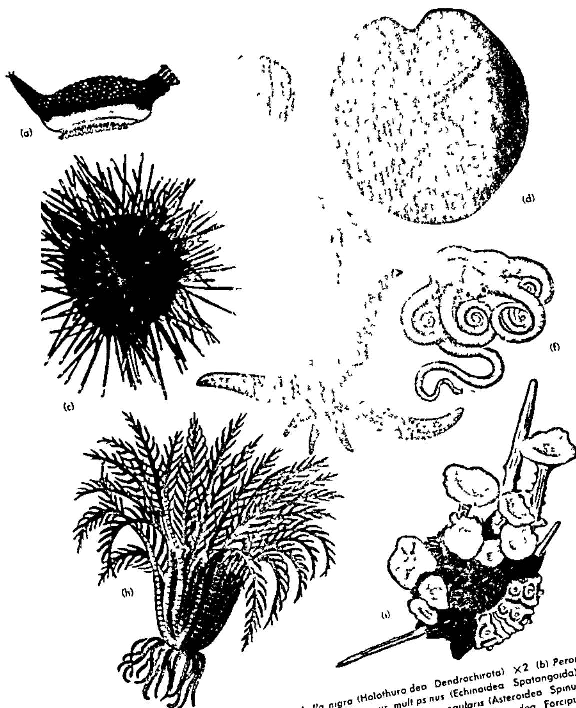


Fig 1 Cross section through radiole of an echinoid showing microscopic structure of skeleton The stereom forms a continuous mesh and the stroma which se cretes it lies in the interspaces which, in this case are arranged in radial rows

Skeleton The skeleton provides the only unique character known from every group of echinoderms from the Cambrian to the present day namely every skeletal plate is composed of a single calcite crystal This feature has facilitated the interpreta tion of some enigmatic Paleozoic fossils which show no other obvious key characters (see CAR POIDEA) During development each plate grows from a group of calcite spicules which are secreted by living tissue termed stroma and fuse together in crystalline continuity Thus the microscopic structure of a plate takes the form of a lattice the stereom with the stroma occupying the meshes It follows that the whole plate exhibits the optical



Representative echinoderms (a) Psol d ella nigra (Holothuro dea Dendrochirota) ×2 (b) Peronella h nemoae (Echinoidea Clypeasteroida (c) Spatangus mult ps nus (Echinoidea Spatangoida) (d) Pseudochinus flemingi (Echinoidea Temnopleuro da) (e) Asterina regularis (Asteroidea Spinulosa) (f) Astrethorax waitei (Ophiuroidea Euryalse) (g) Allostichaster ins gnia (Asteroidea Forcipulata) ×2 (h) Comanthus benhami (Crinoidea Art culata) (i) Goniocidaris parasol (Echinoidea Cidaroida) (b c d i) The reds and purples are mainly the result of echinochromes (a h) Melano ds and carotenoids are present (e f g) The green yellow and orange pigments are carotenoids combined with proteins



properties of mineral calcite (see CALCITE) Although the surfaces of a plate do not usually coincide with the faces of the crystal composing it its major dimension usually matches one of the three molecular axes of the crystal In fossils the spaces in the mesh often become filled by secondary infiltrated calcite and in such cases the plate exhibits the rhombohedral cleavage of calcite The arrangement of the plates differs in the included classes They tend to form symmetrical rows either radiate or bilateral but in holothurians they may remain as scattered spicules

Symmetry Symmetry of the radiate type which characterizes modern echinoderms and many fossils was not an original feature of the phylum Carpoids show no radial symmetry at all and other groups show a blending of radial and bilateral symmetry Larval stages are bilaterally symmetrical When radial symmetry is predominant it is typically fivefold termed pentamerous Radiate structure most constantly shows in the arrangement of the tube feet These usually lie in five equally spaced spokes the radii around the mouth The intervening spaces are termed *interradii* If the radii grow out beyond the interradii the body becomes star shaped (Asterozoa) If the radii develop lateral side branches the arms resemble fern fronds and the body has a flowerlike form (Crinoidea) If the radii and interradii remain in contact and no arms develop the body is more or less globular (Echinoidea) or cylindrical (Holothuroidea)

Alimentary system This system is typically a digestive canal lined by endoderm running from the mouth to the anus The included groups differ widely as to the position of the mouth the presence or absence of an anus and of digestive caeca and other details Two major groups (subphyla) are distinguishable in one the mouth is directed upward (see PELMATOZOA) in the other downward (see ELEUTHEROZOA)

Reproductive system The reproductive organs comprise interradial gonads with separate ducts to the exterior except in *crinoids* in which the gonads lie on the arms The sexes are usually distinct

Coelom The coelom is well developed In the embryo it usually arises as enterocoelous pouches from the gut but in species with direct development a schizocoele forms There are two main portions the perivisceral coelom which surrounds the viscera and contains watery coelomic fluid and the hydrocoele an isolated portion of the coelom from which the water vascular system develops

Water vascular system This is an internal closed system of reservoirs and ducts containing a watery fluid The fluid is mainly drawn from the surrounding sea water sucked in through a delicately perforated sieve plate the madreporite Its minute pores lead into delicate tubes which converge to form one ciliated stone canal (so called because its wall is more or less calcified) The stone canal in turn opens into a circular duct around the mouth the ring vessel From the ring vessel a radial water vessel traverses each of the five radii Tentaclelike outpushings of the radial vessels protrude through the skin These are the tube feet or podia each one often furnished with a reservoir the ampulla

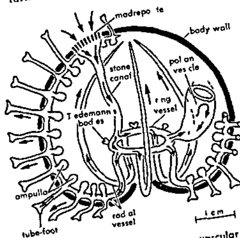


Fig 2 Schemat c d agram of water vascular system in an echino d Arrows show d rection of water flow

sometimes there may be a terminal suction cup operated by a small muscle which raises the roof of the sucker An active echinoderm extends its tube feet by muscular contraction of some other part of the closed water vascular system thus forcing water into the tube foot Suctorial tube feet can adhere to hard surfaces drag the animal along anchor it seize hold of prey and pass food to the mouth In other cases when the tube feet lack suckers they serve as stiltlike limbs (some asteroids) as tentacles (ophiuroids) or they waft food particles toward the mouth (crinoids) Large reservoirs polian vesicles are often associated with the ring vessel The ampullae are subsidiary reservoirs which retract the tube feet by withdrawing fluid from the latter a valve prevents the back flow of fluid from a tube foot to its radial vessel Wandering cells amebocytes float in the fluid and are produced by Tiedemann s bodies small glands opening into the ring vessel In holothurians the madreporite may be internal and accept coelomic fluid instead of sea water In *crinoids* the madreporite may be replaced by scattered hydropores (see CRINOIDEA HOLOTHUROIDEA) In most echinoderms the radial vessels and tube feet are associated with a definite series of skeletal plates such a series is called an ambulacrum In most echinoderms the madreporite is interradial in position and in such cases the interradius in which it lies together with the two adjoining ambulacra constitutes the bivium and the remainder of the body the trivium

Nervous system The nervous system is largely diffuse and often retains its connection with the ectoderm One or several nerve rings surround the mouth send radial nerves into the radii and fol

low a pattern similar to that of the water vascular system

Physiology No definite excretory organs have been identified Specialized respiratory organs occur in the extant Eleutherozoa

Biochromes These are organic pigments which occur in all echinoderms D L Fox (1953) defined *three principal* groups (1) echinochromes (including spinochromes) purple red or green naphthoquinones known only from sea urchins and certain homopterous insects (2) dark melanoids *which occur in* ophiuroids sea urchins and holothurians and (3) carotenoids *which occur in the* integument of all echinoderms They are typically red or orange but if conjugated with a protein they may be *green blue or purple* Alcohol denaturizes the protein so that the carotenoid reverts to red or orange

Phosphagens Phosphagens of two types occur in echinoderms according to E Baldwin and W H Yudkin (1950) These are arginine phosphate in crinoids asteroids, and holothurians creatinine phosphate in ophiuroids and both types in sea urchins Creatinine phosphate *is also* reported from hemichordates and vertebrates whereas arginine phosphate is known from mollusks and arthropods L Hyman (1955) accepts these results as supporting the theory that echinoderms are related to chordates H B Fell (1948) considers that biochemical evidence is inconclusive, not only because it conflicts with paleontological evidence but also because it is self contradictory Echinochromes are found in sea urchins and in some insects, but not in other echinoderms, sterols of one type occur in sea urchins and ophiuroids and a similar type in crinoids, whereas sterols of a different type occur in asteroids and holothurians E Marcus (1958) points out that it is impossible that ophiuroids and echinoids could be more closely related to vertebrates than to the other Eleutherozoa although this would be the implication of the occurrence of creatinine phosphate

Embryology. Surviving Pelmatozoa (crinoids only) have what is essentially a direct development sometimes with a simple yolk larva the vitellaria, which does not feed Some Eleutherozoa have a similar development and yolk larva Fell (1948) suggested that because the vitellaria is common to

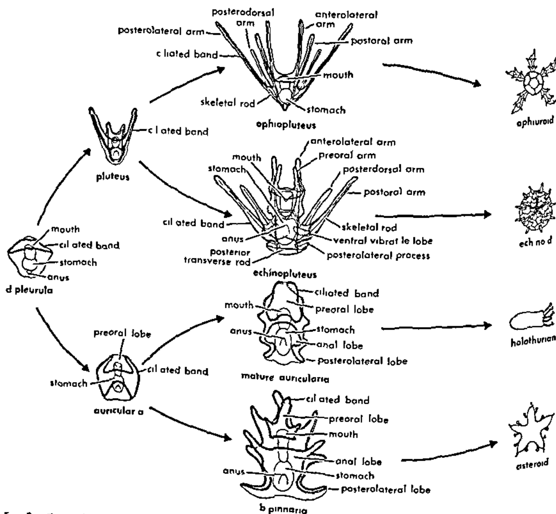


Fig 3 General scheme indicating relationships of bilaterally symmetrical echinoderm larvae

both subphyla direct development (with only a yolk larva) may have occurred in ancestral echinoderms and the coelom may not always have been enterocoelous Hyman (1955) rejects this suggestion as based only on modified types of embryos F Marcus (1958) also maintains that indirect development must be prototypical for echinoderms and protochordates even though most echinoderms may now have direct development

Among those Eleutherozoa which have an indirect development (those with a prolonged food gathering larval stage) two well marked larval types occur (1) the pluteus group with long armed bilaterally symmetrical easel shaped forms common to ophiuroids and echinoids and (2) the auricularia group which are barrel-shaped forms with a winding ciliated band which may be produced into lobes The latter group is common to holothurians and asteroids In most asteroids the auricularia stage is followed by a similar but somewhat more complex larva the bipinnaria or sometimes also by an anchored larva the brachiolaria The chief features of development are illustrated in Fig 3 *See* INVERTEBRATE EMBRYOLOGY

Phylogeny The auricularia larva presents close and striking resemblances to the tornaria larva of some enteropneusts and the enterocoelous development parallels that found in primitive chordates Hence echinoderms and chordates have long been regarded as related groups This well-established theory is now in dispute

The significance of similarities in the larvae of echinoderms and protochordates may be viewed in the following context If the echinoderms are arranged to express their inferred relationships on the basis of their larvae the result places the ophiuroids near the echinoids and apart from the asteroids, which again are placed near the holothurians But this result is in total disagreement with the results of paleontology and morphology both of which indicate that ophiuroids and asteroids are closely related to each other (*see* ASTEROZOA) In addition the paleontology of echinoids is at least as well known as that of any other group of animals and it indicates that echinoids have followed an entirely independent development since the Cambrian On the other hand ophiuroids and asteroids share common post Cambrian ancestors Therefore the remarkable resemblance between the larvae of ophiuroids and those of echinoids can be due only to convergent larval evolution Similarly the differences between the larvae of ophiuroids and asteroids can be due only to larval divergence for both groups arose from Somasteroidea It therefore follows that within the phylum larval similarities do not indicate phylogenetic affinities It is inadvisable to try to extrapolate beyond the phylum so as to infer phylogenetic affinity between hemichordates and echinoderms solely on the ground that the auricularia resembles the tornaria The foregoing analysis was put forward in detail by H B Fell (1948) and has been accepted and supported by N J Berrill (1955) F Marcus (1958) although differing from Fell in believing that indirect development must be prototypical for echinoderms and protochordates agrees that the asteroids and ophiuroids must be closely related and that therefore broad phylogenetic conclusions cannot be drawn on the basis of their larvae Marcus considers that any theory implying that some groups of Eleutherozoa are more closely related to enteropneusts and vertebrates than the four classes of Eleutherozoa are to one another is necessarily absurd L Hyman (1955) who does not discuss the extinct Asterozoa which link ophiuroids and asteroids groups the extant Eleutherozoa as their larval similarities suggest and concludes that the arrangement adopted by paleontologists must be somehow wrong and that the objections of Fell cannot carry any weight J Z Young (1950) likewise believes that the safest evidence of affinity is a similarity of developmental processes and considers that the resemblance of the tornaria to the auricularia is a sure demonstration that chordates are related to echinoderms Fell (1959) reaffirms that the paleontological evidence overrules embryological considerations and maintains his original argument The problem which is fundamental can only be solved by future research *See* ECHINODERMATA FOSSILS

[H B F]

Bibliography F A Bather F S Goodrich and J W Gregory *The Echinoderma* 1900 N J Berrill *The Origin of Vertebrates* 1955 L Cuenot and C Davydov *Echinodermes* in P P Grasse (ed) *Traite de Zoologie* vol 9 1948 H B Fell Echinoderm embryology and the origin of chordates *Biol Rev* 23(1) 81 107 1948 L H Hyman *The Invertebrates Echinodermata* vol 4, 1955 E Marcus On the evolution of animal phyla *Quart Rev Biol* 33(1) 24-58 1958

Echinodermata fossils

An especially important group of fossil invertebrates comprising representatives of the phylum named Echinodermata or spiny skin animals Except for reworked fragments found in some nonmarine deposits echinoderm remains occur exclusively in strata laid down on sea bottoms—chiefly those of shallow seas such as repeatedly submerged large parts of continents during the geologic past Many of these deposits ranging in age from Cambrian to Neogene (Late Tertiary) contain abundant echinoderm fossils of varied sorts Indeed some rocks are composed almost wholly of them In addition echinoderms undoubtedly are widely distributed beneath all oceans although remains of these organisms in deep water sediments even of Recent origin, are virtually unknown because of their inaccessibility Modern echinoderms have been dredged from depths as great as 10 710 meters (35 120 ft) in the Mariana Trench of the southwestern Pacific and are known to be common on the ocean floor at average depths of 4000 meters (13 000 ft) In Fig 1 an ophiuroid echinoderm is shown on the floor of the Ro-



Fig 1 Photograph taken in the Romanche Trench at a depth of 25 000 ft Lat 0°10 S Long 18°21'W Note starfish and angular shaped rocks of several colors Numerous small white objects are probably living since they have a shadow (noticeable on original photograph) showing that they are above the bottom Dark areas caused by mud on camera window (Photograph by H E Edgerton)

manche Trench It has not been possible however to explore ancient deposits formed in deep open ocean waters See ECHINODERMATA

The echinoderms are very well adapted to preservation as fossils owing to their abundant secretion of *calcareous hard parts* Paleontological importance of the group is explained partly by this fitness but more by the *diversity* of their kinds the generally short lived existence and wide geographic distribution of most recognized taxonomic units and the *clearness* with which evolutionary trends generally can be defined Variations are virtually limitless because *no other type* of invertebrate animal outranks echinoderms in complexity of skeletal *structure* which may include upward of 1 000 000 components in an individual

General morphology A *primary* distinguishing feature of echinoderms is the crystalline nature of their skeleton Each *discrete element* consists of calcium carbonate with molecules not distributed at random or as is most common in shells grouped to form fibrous *structure but* arranged in the space lattice of a crystal of calcite All skeletal parts (plates *spines other* ossicles) have a honeycomb structure with soft organic tissue interspersed in the calcareous skeleton yet each part is a crystallographic unit *During* fossilization the soft tissue generally is replaced by calcite so as to make the plate or spine solid

A second very clearly defined character in most echinoderm groups is *pentameral* arrangement of the radially disposed parts of the body and its extensions These function especially for food gathering and in some forms for locomotion This is expressed typically by the five outspread rays of starfishes and the slender arms of brittle stars diverging symmetrically from the central discoid body (Fig 2)

Thirdly deserving notice is a less evident but very prevalent bilateral symmetry in the skeletal structure of echinoderms This serves for orientation of specimens in description supplies evidence for recognition of evolutionary trends in different stocks and guides studies of homology (similarity of structure) in the very divergent main groups of echinoderms

A fourth feature is separateness in development and growth of all the multitudinous skeletal parts They individually increase in bulk as the echinoderm grows Plates of the test covering the body mostly are joined together firmly along sutures but they may fit loosely together or may be separated widely as in holothuroids

Classification Most fossil and living echinoderms are divisible broadly into three groups ranked as subphyla The two main groups are (1) bottom dwelling attached forms collectively known as Pelmatozoa (stemmed animals) that typically are anchored by means of a more or less elongate stalk composed of ligament joined calcareous plates of circular elliptical or pentagonal outline one placed above another so as to make a somewhat flexible column and (2) unattached forms capable of crawling about designated Eleutherozoa (free moving animals)

Pelmatozoa The pelmatozoans evidently include kinds of echinoderms least modified from the very ancient progenitors (surely Precambrian) of the whole assemblage They belong mainly to paleontology A characteristic pelmatozoan feature is orientation of the body with the oral (ventral) side directed upward and aboral (dorsal) side downward although commonly used the designations ventral and dorsal are not aptly suited to these echinoderms (Figs 2c 3) Only a modest remnant of one group the Crinoidea (sea lilies and feather stars) persists as living pelmatozoans whereas extinct classes include the Cystoidea (irregular saclike forms), Blastoidea (budlike) Edrioasteroidea (sessile starfishes) and lesser assemblages called Eocrinoidea and Paracrinoidea None of the five last named groups persisted beyond Paleozoic time

Eleutherozoa The eleutherozoans include six classes most of which are more abundant and varied in modern faunas than in those known during any part of the geologic past (1) The Echinoidea are characterized by a rigid skeleton of rounded form covered by multitudinous long or
s — is promi
s and well
not at all
evident in the extremely numerous globose types (Fig 2a b) (2) The Asteroidea (starfishes) are distinguished by strongly developed radial armlike extensions from the central body and flexibility of the rays (Fig 2d e) (3) The Ophiuroidea (brittle stars) resemble asteroids but are readily distinguished by sharpness of separation of their long slender arms from the discoid body

They are by far the most active echinoderm animals being able to crawl rapidly Formerly the asteroids and ophiuroids commonly were grouped together under the name Asterozoa or Stelleroidea This classification now is discarded in the light of important morphological and embryological evidence indicating wide phylogenetic divergence (4) The Somasteroidea (large-body starfishes) are a primitive early Paleozoic group standing apart from other eleutherozoans (5) The Ophiocistioidea are echinoidlike forms of Paleozoic age In all of the classes just enumerated the mouth is directed downward and the aboral side upward Their orientation thus is opposite to that of pelmatozoans (6) The remaining major group of eleutherozoan echinoderms, designated Holothuroidea (sea cucumbers), is characterized by an elongate cylindrical form and skeletal parts consisting only of discrete, generally microscopic ossicles that occur mainly in the leathery highly flexible body cover The mouth is located at one end of the body and the anus at the opposite extremity (Fig 2*f*, *g*) The animal lies on one of its sides which is constantly downward oriented and accordingly may be termed ventral

Homalozoa An additional echinoderm subphylum small in numbers but otherwise significant is named Homalozoa (flat animals) represented by fossils ranging in age from Cambrian to Devonian These lack a radial arrangement of skeletal parts which consist of irregular plates of crystalline calcite enclosing a bilaterally flattened body Main homalozoan assemblages called Carpoidea and Machaeridia are recognized

Cystoids The class named Cystoidea contains extinct pelmatozoans of globose subcylindrical or somewhat flattened ellipsoidal form that mostly are characterized by irregularity of the plates surrounding the body (Fig 3*a*) They are most common in Ordovician and Silurian marine deposits but are known to range from Early Cambrian to Late Devonian A short generally weak stem furnished attachment and a variable number of slender appendages (termed brachioles) on the upper side of the plated body (calyx) supplemented by ambulacral grooves on this surface served for gathering food The mouth was located centrally at the summit of the calyx and an anus fairly well down on one of the sides which accordingly is defined as posterior Other small orifices interpreted as hydropore and gonopore may also be found near the anus

Cystoids are grouped in two main assemblages on the basis of their plate structures In many genera the plates are perforated by numerous minute tubular openings that extend from outer to inner surface of the plates so as to allow sea water to circulate through them Since the openings typically occur in pairs these genera are known as Diploporita (twin pore forms) (Fig 3c) Such cystoids have exceptionally numerous calyx plates that lack any perceptible regularity of arrangement Remaining cystoids are named Rhombifera (rhomb bearing forms) because some or all of

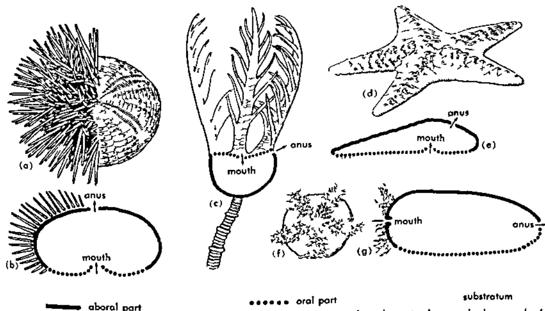


Fig 2 Representative types of echinoderms (a) Regular echinoid (*Lytechinus* Recent) oblique aboral view right half with spines removed showing two ambulacra (narrow petal like tracts) and three interambulacra (b) Diagrammatic section through a showing stoutly built test (c) Crinoid showing stem and three arms calyx sectioned with parts distinguished as in b (d) Asteroid (*Dermasterias*, Recent) aboral oblique view (e) Diagrammatic section through ray at left and opposite interambulacrum (f) Holothuroid oral view showing tentacles around mouth and orientation of rays (g) Diagrammatic section of holothuroid

their calyx plates are penetrated by rhomb shaped groups of tubes running parallel to each other and to the plate surfaces (Fig 3a) The two halves of any rhomb lie on adjoining plates with the tubes crossing the plate boundaries at right angles (Fig 3b) The function of the tubes, which open externally at their ends, presumably was respiratory The rhombifers possess relatively few calyx plates and these exhibit fair constancy of arrange

calyx plates suggests that the blastoids may be descendants of rhombiferan cystoids

Blastoids. An unusually stable class of pelmatozoans consists of the Blastoidea (Fig 3e) They

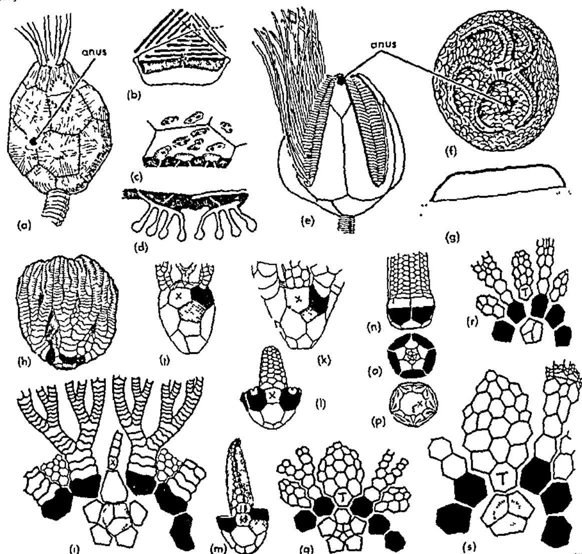


Fig 3 Fossil pelmatozoan echinoderms (a) Rhombiferan cystoid (*Echinoencrinus* Ordovician), showing pore rhomb and anus on side of calyx (b) Diagrammatic oblique view and section through cystoid pore rhomb (half of one rhomb toward front bisected by suture between plates) (c) Oblique view and section of diploporitan cystoid plate (d) Section through ambulacrum of blastoid (*Pentremites* Mississippian) showing hydrospires beneath lancet and side plates (e) Blastoid (*Pentremites*) with brachioles restored along side of one ambulacrum, showing deeply forked radial plates succeeded above by diamond-shaped deltoids (f) Edrioasteroid (*Carneyella*, Ord), oral view, showing strongly curved ambulacra diverging from slitlike mouth (g) Section through stalkless edrioasteroid (h, i) Flexible crinoid (*Taxocrinus* Miss) posterior side and partial plate diagram (radials black) showing right posterior plane of bilateral symmetry defined by infrabasal circlet (j p) Inadunate crinoids some showing anal sacs and parts of arms j *Carabocrinus* Ord k, *Botryocrinus*, Devonian, l, *Cyathocrinites*, Miss m, *Cupulocrinus*, Ord, n p, *Delocrinus*, Pennsylvanian (q) Diplobathrid camerate crinoid (*Ptychocrinus*, Ord), plate diagram (r, s) Monobathrid camerates r, *Macrostylocrinus*, Dev, s, *Periechocrinites*, Miss, plate diagrams showing noteworthy distinctions in basal and radial circlets

display a prominently developed, very regular fivefold radial symmetry, on which are superposed a first order of bilateral symmetry in the anteroposterior plane and a second order directed through the left posterior ray (Fig 4b) The calyx of average-size specimens is small, with diameter of about 15 mm and height of 20 mm It is composed of 23 or 24 plates of which 5 (lancet plates) are not visible externally on unweathered specimens From the circular stem facet upward, the visible calyx plates include 3 basals (2 large and 1 small), 5 deeply forked radials, and 5 generally small diamond-shaped deltoids in interradial position, in several genera the posterior interradius contains 2 deltoid plates epideltoid above the anus and hypodeltoid below it Within forks of the radials are a multitude of very diminutive so-called side plates that conceal the lancets, they form ambulacral areas bordered laterally by rows of threadlike free armlets (brachioles) that function as food gatherers Particles of food are conducted to midlines of the ambulacra and thence upward to the mouth, which is at the center of the ventral surface Beneath the ambulacra are extremely delicate longitudinally folded calcareous lamellae that enclose narrow troughs for circulation of water admitted through pores at the base of the brachioles, these structures, termed hydrospires, correspond to the slit like parts of pore rhombs in the rhombifer cystoids (Fig 3d)

The oldest known blastoids occur in Middle Ordovician rocks They are more numerous but not common as fossils in Silurian and Devonian formations, attain extraordinary abundance in some Mississippian strata, and then virtually disappear until a brief burst is recorded in Lower Permian rocks, especially in the island of Timor, East Indies

Edrioasteroids A unique assemblage of mainly early Paleozoic attached echinoderms is named Edrioasteroidea (Fig 3f g) These fossils, distributed from Lower Cambrian to Upper Pennsylvanian have no stem for anchorage but were fixed by their entire base The upper (ventral) side was covered by many flexibly joined plates, which are sharply differentiated into rows of paired ambulacrals with adjoining adambulacrals and ir

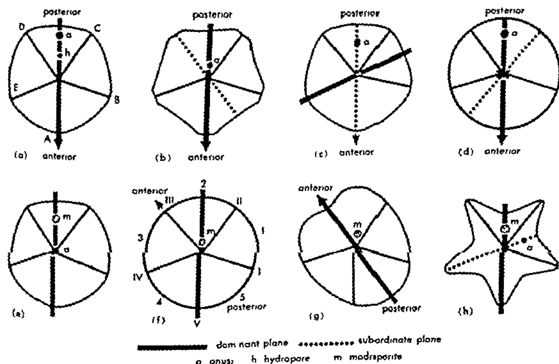


Fig 4 Bilateral symmetry developed in various echinoderms Anterior and posterior directions are indicated where distinguished All diagrams represent aboral views. (a) Pelmatozoans generally (most crinoids, cystoids, edrioasteroids) letters denote ray designations according to the Carpenter (1884) system A anterior B right anterior C right posterior D, left posterior, E, left anterior (b) Blastoids, heterocrinoids, with subordinate bilateral symmetry in right posterior plane (c) Homocrinoids, with primary bilateral symmetry in left anterior plane (d) Glyptocrinoids, flexible crinoids, and rhombiferan cystoids, with subordinate bilateral symmetry in right posterior plane (e) Holothuroids (f) Regular echinoids rays marked according to the Lovén (1874) system, III being considered anterior (g) Irregular echinoids, with prominent bilateral symmetry in left posterior plane of pelmatozoans (h) Asteroid with subordinate bilateral symmetry in left anterior plane

regular interambulacrals The five ambulacral tracts curve outward from the centrally located mouth with their extremities typically deflected rather strongly in a consistent way The anus is located in one of the interrays that accordingly is denominated as posterior

Eocrinoids A small group of Cambrian and Ordovician pelmatozoans that combines characteristics of cystoids and crinoids yet differs significantly from both is distinguished as Eocrinoidea They resemble cystoids in mode of branching of the ambulacral grooves and ventrolateral location of the anus but lack thecal pores or distinct pore rhombs they are like crinoids in plate structure and identity or near identity of plate arrangement in the calyx The eocrinoids have stems and unbranched or bifurcating arms

Paracrinoids Paracrinoidea consist of stem bearing echinoderms that also combine features of cystoids and crinoids but in manner quite unlike that of the eocrinoids The paracrinoids which now are known only from Middle Ordovician deposits have numerous plates of the calyx which are not arranged in series and which lack a ventrally differentiated area corresponding to the tegmen of crinoids The plates have a cystoidlike pore structure but the arms are comparable to those of crinoids

Crinoids By far the most important division of pelmatozoans is the Crinoidea Indeed in abundance of fossil remains this group considerably outranks all other echinoderms combined (Fig 2c) It is represented by upward of 5000 extinct species and at least 630 living kinds Some Paleozoic deposits evidence existence of uncounted trillions of these animals They exhibit an amazing variety of forms classifiable in four main groups designated Inadunata Flexibilia Camerata and Articulata Except for a single small family of inadunates of Triassic age the first three of these are restricted to Paleozoic time whereas the fourth includes virtually all post Paleozoic forms

Skeletal features Although many modern and some ancient crinoids are stemless as adults this group of pelmatozoans typically is attached to the sea bottom by a more or less *elongate* stem com[illegible] [illegible] of rootlike branches *embedded* in sediment or a discoid expansion that *may* be cemented to some foreign surface *At the* opposite extremity of the stem which *exceptionally* may be 50 ft tall is the crinoid body encased in regularly arranged plates and surmounted by branched or unbranched free moving *arms The conjoined* plates below the free arms comprise the so-called dorsal cup this cup along with plates of the ventral surface comprising the tegmen makes up the crinoid calyx (Fig 3*h–s*) The calyx and its attached arms are termed the crown The mouth is located on the ventral surface or beneath it. The anus may be placed also on the summit of the calyx with or without considerable elevation above the general level of the tegmen or it may be found on the side of the dorsal cup The posterior side of the crinoid is defined by position of the anus in one of the interrays or by extra plates introduced in such position on one side of the dorsal cup the ray opposite to the posterior interray is defined as anterior This establishes a plane of bilateral symmetry that more or less strongly modifies the fundamental radial symmetry of the crinoid (Fig 4*a*)

The plates of the crinoid dorsal cup are arranged in a regular pattern of successive circlets but with differences in various groups that form a chief basis for classification Each circlet normally contains five plates At the base of each ray is a plate termed radial Beneath the circlet of radials are five basals that are offset in interradial position some crinoids have a still lower circlet of infrabasals that alternate with the basals and hence occur in radial position Crinoids with a single circlet of plates below the radials are termed monocyclic (Fig 3*r*, *s*) and those with two circlets dicyclic (Fig 3*h–q*) Distinction based on this character has prime importance in classification The posterior interradial side of the cup commonly contains one or more extra plates (in different types of crinoids) including *plates* named anal or "X" radianal and tergal or "T"

In some crinoids the tegmen located between arm bases on the ventral side of the calyx is stoutly constructed of small irregularly arranged plates whereas the summit of others *consists of a* flexible leathery integument that may *be studded* with calcareous ossicles Five *subequal plates* larger than others of the tegmen may *occur* interradially located around the mouth *these are* orals A cylindrical globose or umbrella-shaped extension of the tegmen upward (anal sac) is common in some crinoids (Fig 3*k–m*)

The arms of crinoids are extremely variable in plan and construction although *essentially* constant within each genus They may be unbranched or moderately to highly branched with or without very numerous tiny branchlets called pinnules and composed of a single or double series *of* arm plates (brachials) These characters *along* with the mode of articulation among plates of the rays are important also for classification

Main types Three main groups (subclasses) of Paleozoic crinoids are recognized each of them distributed from Ordovician to Permian

1 The Inadunata (not united referring *to* lack of incorporation of lower arm plates in dorsal cup) are crinoids with a relatively small *dorsal* cup containing either one or two circlets *of* plates below the radials and having the arms *entirely* free above the cup (*Fig* 3*j–p*) They include

1750 or more described species that exhibit ut most difference in form and evolutionary trends In a majority the anteroposterior plane of bi lateral symmetry is well marked without other deviation from a regular pentameral plan (Fig 4a) but in one group (superfamily Homocrinoidae) a surprising degree of bilateral symmetry was de veloped in the plane of the left anterior ray (Fig 4c) Some of these crinoids have a strongly downbent crown that was hinged on the summit of the stem In another group (superfamily Heterocrinoidae) a subordinate plane of bilateral sym metry is oriented in the left posterior plane (Fig 4b) The inadunates are classified in three orders Disparida (Ordovician Permian) 34 gen era Hybocrinida (Ordovician) 6 genera Cladida (Ordovician Triassic), 230 genera The first two are monocyclic and the third dicyclic

2 The Flexibilia (flexibles) are a distinctive assemblage of exclusively dicyclic crinoids char acterized by movable ligamentous union between most of the plates and several constant features in organization of the calyx (Fig 3h i) They in clude approximately 300 described species all of which exhibit a secondary bilateral symmetry in the right posterior plane in addition to their gen erally well marked primary bilateral symmetry directed anteroposteriorly (Fig 4d) Flexible cri noids are unequally divided in two orders Taxocrinida (Ordovician Mississippian) 6 genera Sa genocrinida (Silurian Permian) 44 genera

3 The Camerata (chamber or 'box forms) are most numerous among ancient crinoids both in variety and quantity of individuals (Fig 3q s) Some 2500 species of these fossils have been de scribed of which 1650 come from Mississippian rocks alone The camerates are distinguished by the stout construction of their calyx which in corporates lower ray plates and interradials in the dorsal cup and subtegminal location of the mouth Types with both one and two circlets of plates beneath the radials are common Anteroposterior bilateral symmetry is developed almost universally in these crinoids as modification of the dominant pentameral pattern (Fig 4a) in addi tion a secondary plane of bilateral symmetry di rected through the right posterior ray prevails in the monobathrid suborder Glyptocrinina as in all flexible crinoids (Fig 4d) Camerate crinoids are classified in two orders Diplobathrida (Ordovi cian Mississippian) 34 genera dicyclic Mono bathrida (Ordovician Permian) 120 genera monocyclic

4 Mesozoic and Cenozoic crinoids belong to the subclass Articulata represented by about 500 spe cies of described fossils in addition to equally numerous living kinds A majority of this group are stemless as adults and tend to have nearly perfect pentameral symmetry

Echinoids The class Echinoidea is decidedly the most important group of fossil eleutherozoans (Figs 2a b and 5) Like starfishes and others, they are distinguished from pelmatozoans by free dom from a sessile mode of life and except for holothuroids by downward orientation of their oral surface Seemingly paleontological preemi nence of the echinoids is explained mostly if not entirely by the prevailingly rigid construction of their calcareous skeleton for the hard parts of other eleutherozoans tend to be loosely joined or wholly disconnected in manner that is ill suited to preservation showing skeletal organization of the whole animal Modern species of echinoids are approximately 750 in total number as shown by T Mortensen s recently published (1928–1957) several volume monograph covering all known liv ing kinds In comparison with this starfishes amount in the aggregate to some 2000 species and brittle stars at least 1600 species Present day holothuroids include 500 species All eleu therozoan groups are known as far back as Or dovician time and yet fossil echinoids are con siderably more numerous than all others of the subphylum combined Such disparity reflects in equality in fitness of the different sorts of skeletons to be preserved However it is true that not all parts of echinoid skeletons generally are found intact Multitudes of movable spines attached to the test during life tend to be separated and scattered as does the echinoid masticatory appara tus which is known as Aristotle s lantern Thus knowledge of the entire skeletal structure of many fossil echinoids is incomplete

Morphology The echinoid test is typically glob ular unevenly ovoid or discoid in shape (Figs 2 5) Mainly it is formed by 10 meridionally dis

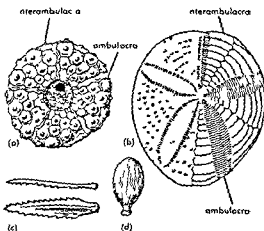


Fig 5 Fossil echinoids (a) Regular form (Echinocrinus Mississippian) aboral view showing anus and periproct wide interambulacra (ia) and narrow ambulacra (a) (b) Irregular form (Eupatagus Paleogene) aboral view left side showing unweathered appearance with many spine bases, right side showing plate structure with ornamentation omitted anterior ray directed upward (c d) Types of echinoid spines c Porocidaris Paleogene d Balanocidaris Jurassic

posed bands of plates of which 5 are ambulacral distinguished by porelike openings between or through the plates the 5 others that alternate with them are interambulacral and lack perfora tions In most echinoids both living and fossil the plates are joined together rigidly but in a mi nority the union is somewhat flexible and imbri cating Virtually all post Paleozoic echinoids are distinguished by a constant arrangement of the ambulacral and interambulacral groups of plates each of the 10 meridional bands comprising a double column making 20 columns of plates in the test as a whole Naturally the plates are widest at the equator (ambitus) of the test and increasingly narrow toward the oral and aboral poles The mouth located on the underside is sur rounded by an area of naked integument or flexi bly united small plates the pattern of which may have taxonomic value the around mouth tract is termed peristome The anus generally is placed on the aboral side of the test opposite the mouth but it may be shifted backward to a lateral posi tion or even to the oral surface as seen in various fossils The anus is surrounded by a bare or small plated periproct Another part of the test impor tant for orientation and classification is the so called oculogenital system of 10 special plates at the aboral pole one of these plates generally somewhat larger than the others is an interambula crally placed

constitutes a

ogous parts

and living ec

erally Study of these homologies is essential in at tacking problems of phylogeny and definition of evolutionary trends

Fossil remains of echinoids include numerous spines that almost invariably are found as isolated *specimens* not associated with the test that bore them Even so they are commonly distinctive and useful as paleontological tools for stratigraphic correlations and age determinations *They are vari* ously shaped some long and slender others short and amazingly bulbous or with spadelike form *surface ornamentation* also may be diagnostic and internal structure helpful for identification (Fig 5c d) Individual plates of echinoid *tests are* com mon fossils in some Paleozoic formations although generally not in younger deposits they are recog nizable mainly by the *rounded* protuberances (mamelons) for spine attachment on their exter nal surface Finally there are isolated *Aristotle's* lanterns or individual teeth and other parts derived from them among fossils these are not sufficiently abundant or well enough known *to merit much* at tention

Orientation Seemingly *there should* be little reason for discussing the orientation of echinoids because the downward directed oral surface clearly is ventral and at least *among irregular* echinoids which are characterized by obvious bi lateral symmetry locomotion consistently is in a single direction along the path defined by the plane of this symmetry If such direction of movement is considered to be forward which rationally seems unavoidable the ambulacral ray on this side of the test must be anterior and the opposite inter ambulacrum is posterior (Fig 4g) In irregular echinoids such as the spatangoids the spines

most universal adoption by special of numerical designation of the rays both ambula cral and interambulacral introduced by S L Loven (1874) and hence known as the Loven sys tem (Fig 4f) It is applied to fossils as well as living echinoids and to both irregular and regular types Orientation of the regular echinoids seem ingly should not be easy since they move in any di rection with equal ease and without detectable preference also except for the madreporite the test has perfect radial symmetry Inclusion of the regulars is possible only by use of the off center location of the madreporite on the aboral surface as a reference point assuming that its relative po sition is the same in all echinoids which is hardly open to question Then the interambulacrum con taining the madreporite is identified as number 2 of the Loven system

Among pelmatozoans which almost certainly embrace the ancestors of echinoids the posterior interradius is readily and positively identified If this orientation is applied to echinoids a discrep ancy becomes evident at once for the adopted an teroposterior plane of echinoids clearly is that coinciding with the left posterior ray of the pelma tozoans (Fig 4e f) Interambu*lacrum 2 (Loven)* considered as right anterior by echinoid students is equivalent to the posterior interambulacrum of crinoids for example This does not mean that echinoid orientation is wrongly conceived but merely that evolution of these echinoderms has pursued a divergent tack of its own which inci dentally duplicates the subordinate plane of bi lateral symmetry in some pelmatozoans (blastoids heterocrinoids)

Classification Taxonomic arrangement of the echinoids generally has recognized two main divi sions treated as subclasses these respectively com prise the regular echinoids (Regularia) and irregu lars (Irregularia) Present knowledge based importantly on fossils as well as living forms indi cates that this arrangement is artificial The so called regulars are a composite assemblage that on one hand contains several extinct primitive kinds exclusively Paleozoic and on the other advanced forms of modern type all Mesozoic and Cenozoic that include the stocks (Diadematacea Echinacea) from which irregular echinoids were derived Ac cordingly classification recently has been consider ably revised (J W Durham and R V Melville 1957) so as to take account both of Mortensen's comprehensive work on living echinoids and ext

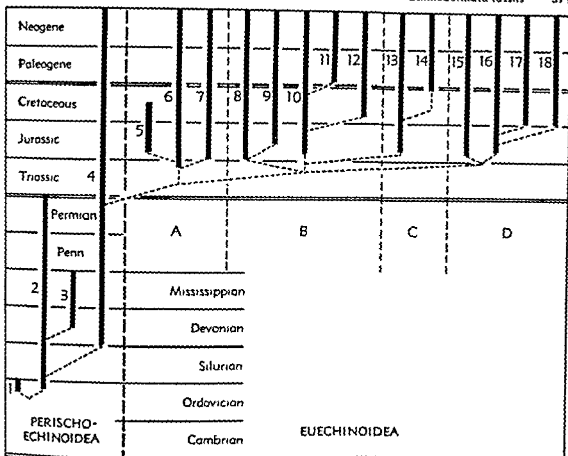


Fig 6 Classification of echinoids showing phylogeny and geologic distribution (based on Durham and Melville 1957) Orders 1, Bothriocidaroida 2, Echinocystitoida, 3, Palaechinoida, 4, Cidaroida, 5 Pygasteroida 6, Echinothurioida 7, Diadematoida 8, Hemicidaroida 9, Arbacioida, 10, Phymosomatoida 11, Echinoida 12 Temnopleuroida 13 Holectypoida 14 Clypeasteroida 15 Holasteroida 16, Nucleolitoida 17 Cassiduloida 18, Spatangoida Superorders A, Diadematacea, B, Echinacea C Gnathostomata D Atelostomata

dence from paleontology In diagrammatic form that shows inferred phylogeny, the new classification is given in Fig 6

Ophiuroids The brittle stars, or Ophiuroidea are treated next because both morphological and embryological evidence (L H Hyman 1955) indicates that these highly mobile echinoderms are more closely related to the echinoids than to star fishes The ophiuroids are distinguished readily by their external form since the rounded to pentagonal or scalloped body is a small flattened disc sharply separate from the symmetrically placed long slender arms that extend radially outward from it The arms may be smooth or spiny Almost invariably they are five in number and unbranched but a few kinds, such as the Recent basket stars (Gorgonocephalidae), show extraordinarily repeated bifurcations, all known fossil ophiuroids have simple arms The skeleton of the central disk is composed of many regularly arranged plates, without any deviation on either the aboral or oral surfaces from perfect pentamerous symmetry The mouth is at the center of the underside but an anus is lacking The skeletal structure of the arms is distinctive in being internal and composed of fused ambulacral ossicles

Fossil ophiuroids are known from Lower Ordovician to Pleistocene (Recent) but they are not abundant Only 50 genera and less than 100 species have been described from Paleozoic, Mesozoic, and pre Recent Cenozoic formations as compared with approximately 1600 known living species As a whole, ophiuroids are classed in two main divisions Stenurida ranging from Lower Ordovician to Lower Devonian and Ophiurida, distributed from Lower Ordovician to Recent All modern ophiuroids belong to the Ophiurida

Asteroids. The starfishes, or Asteroidea, are comparable to the ophiuroids in antiquity of their lineage, as in their general stellate form (Figs 2*d*, *e*, 4*h*) They differ from the brittle stars in two main respects lack of strongly marked separation be

tween the body and its radially disposed extensions and the ventrally open ambulacral grooves along the rays The skeletal elements tend to be either loosely joined or quite separate Hence asteroids are not very well suited for fossilization in manner showing the skeletal arrangement of the entire animal Ossicles along the ambulacra occur in two or four series discriminated as ambulacrals and adambulacrals Some asteroids lack armlike extensions consisting of pentagonal shapes with ambulacral grooves on their underside Others such as the Heliasteridae with as many as 44 rays lack evident pentameral symmetry but structurally resemble common types of asteroids

Classification of the Asteroidea that takes account of both fossil and modern forms generally has been lacking Recent work by W K Spencer and C W Wright covering both fields for the *Treatise on Invertebrate Paleontology* indicates a classification that recognizes four orders (Paxillosida Forcipulatida Spinulosida Valvatida) all of which are represented by Paleozoic to Recent forms

Somasteroids Ranked as correlative with the Asteroidea and Ophiuroidea but represented only by some Lower Ordovician forms are starfishlike fossils named Somasteroidea They differ evidently from other eleutherozoans in the broad petaloid nature of the rays which contain relatively long series of rodlike ossicles diverging outward from ambulacral plates in each ray Ambulacral grooves for transportation of food particles to the mouth are shallow or in forms interpreted as most primitive not recognized The aboral surface had a coarse skeletal meshwork

Holothuroids The sea cucumbers or Holothuroidea are a group of eleutherozoans that must have diverged very early from other echinoderm stocks They clearly exhibit a fivefold radial symmetry characteristic of the phylum but otherwise differ radically from any eleutherozoan or pelmatozoan assemblage They are greatly elongated along the oral aboral axis so as to have the subcylindrical form of link sausage or a cucumber and they lie on one of their sides identified as ventral because it uniformly *is placed downward* (Figs 2*f g 4e*) The mouth surrounded by dendritic tentacles lies at one extremity and the anus at the opposite end The body is encased *in a flexible* somewhat leathery integument that contains microscopic calcareous plates loosely *distributed* When the animal dies decay of the uncalcified tissues liberates the skeletal parts *which become scattered* about Fossil remains of holothuroids thus consist only of discrete microscopic ossicles Even so numerous kinds of these parts *are preserved* in rocks ranging from Ordovician to Pleistocene

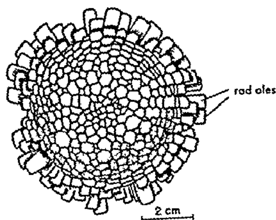
Ophiocistioids An aberrant eleutherozoan group containing five known *genera distributed* from Lower Ordovician to Middle Devonian is named Ophiocistioidea The plate-enclosed body roughly corresponds to that of echinoids but noteworthy differences indicate that the ophiocistioids are at most extremely remote relatives Chief peculiarity of the group is the presence of rather numerous armlike appendages attached to the oral surface in ambulacral position They are interpreted as much enlarged podia that are covered by abundant minute plates

Origin and phylogeny The origin of the echinoderm invertebrates surely belongs some time before the beginning of the Paleozoic Era for Lower Cambrian deposits contain such divergent branches of the phylum as Carpoidea Edrioasteroidea and Eocrinoidea These are primitive sorts of pelmatozoans Cystoids crinoids and blastoids as well as all recognized main groups of eleutherozoans appear in Ordovician strata It is evident from the already well marked development of characteristic structures that a long period of antecedent evolution must have operated to produce such widely divergent forms For example each main division of the Paleozoic crinoids (Inadunata Flexibilia Camerata), although seeming to appear very suddenly is fully differentiated in the Middle Ordovician One can only guess as to reasons for the absence of these echinoderms in older deposits During the Paleozoic numerous well marked evolutionary trends are discernible in nearly all echinoderm groups including eleutherozoans (especially echinoids) as well as pelmatozoans All groups of modern echinoderms have their origin in early Paleozoic stocks and the lines of their phylogenetic are mostly indicated by the fossil record Echinoids predominate in Mesozoic and Cenozoic echinoderms [R C M]

Bibliography J W Durham and R V Melville A classification of echinoids *J Paleontol* 31(1) 242–272 1957 L H Hyman *Echinodermata* vol 4, 1955, R C. Moore Echinodermata Pelmatozoa *Bull Museum Comp Zool* (Harvard) 112(3) 125–149 1954, R C Moore and L. R Laudon *Evolution and Classification of Paleozoic Crinoids* Geol Soc Am Spec Paper 46 1943 R C Moore C G Lalicker and A G Fischer *Invertebrate Fossils* 1952, J Piveteau (ed) *Traité de Paleontologie*, vol 3 1953, R R Shrock and W H Twenhofel *Principles of Invertebrate Paleontology* 2d ed 1953

Echinoida

An order of Echinacea with a camarodont lantern smooth test imperforate noncrenulate tubercles ambulacral plates of echinoid *type* and shallow branchial slits (*see* ECHINOIDEA) There are numerous tropical and temperate species *some* of them remarkably adapted to living on coral *reefs* (see illustration) The four included families are mainly distinguished by characters of the pedicellariae The Parasaleniidae are oblong forms with trigeminate ambulacral plates They range from the Eocene to the present day The Echinidae possess trigeminate or polyporous plates with the pores in a narrow vertical zone Strongylocentrotidae are polyporous with the pores in 2–4 vertical series The Echinometridae show a variety of forms



Colobocentrotus atratus aboral aspect A Pac fc spe c es adapted for l fe on wave-exposed coral reefs

which include polyporous types with an oblong test and trigeminate or polyporous types with a spherical test *See* ECHINACEA [H B F]

Echinoidea

A class of Eleutherozoa known as the sea urchins These animals have a compact body enclosed in a hard shell or test formed from regularly arranged plates which bear movable spines (Fig 1) There are no arms but radii are represented by 5 double rows of tube feet arranged as meridians between the upper and lower poles of the body

There are about 850 living species and some 5000 fossil species have been recorded included in 225 genera They are classified in 18 orders grouped in 2 subclasses Sea urchins range in size from a few millimeters across the test to 20 cm They differ from other echinoderms in possessing echinochromes in the pigmentation complex Although many species are dull or dark in color some are brilliant shades of purple red green or orange Others have particolored striped spines and deep-sea forms may be white

Relation to man Some tropical species have hollow brittle spines which cause septic wounds if they break off after penetrating the skin The genus *Araeosoma* carries venomous spines which can inflict dangerous wounds and a related genus *Asthenosoma* according to T Gislen can kill a man (see ECHINOTHURIOIDA) Most species are quite harmless and many are eaten in various lands where the sex glands are esteemed in season One species (*Tripneustes ventricosus*) is the subject of a legally regulated fishery in Barbados

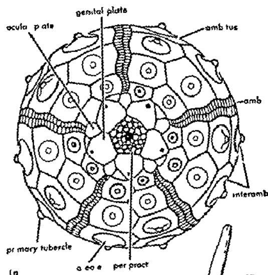
Ecology Sea urchins occur in all seas from low tide level downward *Pourtalesia* reaches a depth of more than 4 miles (7250 m) in the Banda Trench The rounded (or regular) urchins feed mainly on algae often hiding by day under stones which are held over the test by tube-feet and emerging at night or at high tide The heart urchins and other exocyclic forms live buried in mud or sand feeding either on organic matter in the mud or on selected detritus

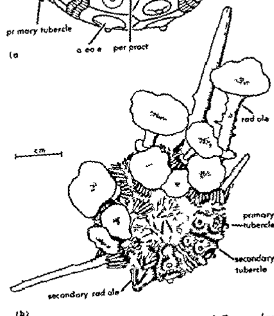
Sea urchins move slowly using muscles at the bases of the spines to swing the spines like stilts The suctorial tube-feet are used to ascend steep surfaces and as anchors

Numerous parasites have been recorded Among these are protozoans nematodes and gastropods of which several genera bore into the test Crabs live in the rectum or on the test feeding on spines and tube-feet Other animals shelter among the spines or attach themselves to exposed hard parts

Skeleton The test is globular or nearly so in those forms in which the anus lies within the apical system (Fig 2*a*) These are often called Regularia In exocyclic forms (Fig 2*b*) the test tends to assume a secondary bilateral symmetry (*see* IRREGULARIA REGULARIA) The radial symmetry is always evident however and is 5-part (pentamerous)

Test The test is built up of regularly arranged plates which collectively form a rigid or sometimes flexible investing shell In all echinoids since the Triassic the test is composed of 10 meridional





F g 1 Structure of the ech no d test of *Goniocidar s parasol* (a) Naked test (b) Test w th radioles

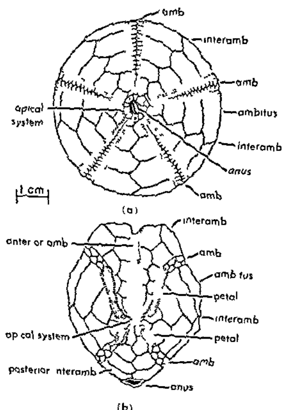


Fig 2 (a) Endocyclic echinoid (b) Exocyclic echinoid

areas each composed of 2 vertical columns of plates The meridians converge above and below at the upper and lower poles The equatorial zone is termed the ambitus Of the 10 meridional areas 5 correspond to ambulacra because each of them carries a double row of tube feet The alternating 5 meridional areas are termed *interambulacra* or interambs They are usually wider than the ambulacra or ambs

Paleontology reveals that the test evolved only after much trial and error because the earlier Paleozoic forms shows extreme instability in the number of columns of plates (see PERISCHOECHINOIDEA) The cidaroids were the first group in which stability was achieved (in the Permian and Triassic) but a rigid spherical shape was not adopted until the Jurassic (see CIDAROIDA) All the bilaterally symmetrical exocyclic forms evolved from cidaroid ancestors

Radioles Most echinoids carry spines or radioles which articulate with tubercles on the test plates and are moved by muscles at the base of the spine Large radioles (primaries) articulate with large primary tubercles smaller ones with secondary tubercles If a ligament links the radiole socket to the tubercle the tubercle has a small hole in it where the ligament is attached and is termed perforate Tubercles without such a ligament are imperforate The radiole muscle may impress an indented pattern on the edge of the boss on which the tubercle stands, if so the tubercle is termed crenulate if not noncrenulate The muscle is attached to the test plate on a saucer shaped depression around the tubercle the areole All these features are used in systematic diagnoses of the orders and families

At the upper pole lies a circlet of 10 plates the apical system Of these 5 top the ambs and the 5 which alternate with them top the interambs Those which top the interambs each carry a gonopore and are termed genital plates One of them also acts as a madreporite The other 5 are termed ocular plates although each carries a tentacle not an eye In the Paleozoic genus *Bothriocidaris* one of the oculars served as the madreporite which was therefore radial in position a unique feature In the spherical echinoids (Fig 2) the anus lies within the apical system on a membrane termed the periproct These forms are termed *endocyclic* If the anus becomes displaced outside the apical system it enters as interradius (interamb) termed the posterior interamb and the echinoid is said to have become exocyclic These features were formerly used in classification and are still valuable in determining trends of development in evolution Exocyclic forms may exhibit modification of the apical system The anus, for example tends to obliterate the posterior genital plate (and its gonad) as it migrates backward and the lost structures may not be replaced or considerable distortion may ensue See CASSIDULOIDA CLYPEASTEROIDA HOLASTEROIDA

Lantern The mouth lies on the lower (oral) side of the test surrounded by soft skin the *peristome* In endocyclic forms the mouth is central but in exocyclic forms it may suffer displacement into the radius and lie opposite the interamb which contains the anus The radius is then termed *anterior* In endocyclic forms and in some of the exocyclic forms the mouth is furnished with a ring of five powerful jaws, each with one large tooth The jaws and teeth collectively comprise the so-called *lantern* first described by Aristotle The lantern is moved by muscles which are attached to an inturned flange of the test around the peristome the perignathic girdle The structure of the girdle lantern and teeth are characteristics used in classification The teeth may be grooved or keeled (Fig 3)

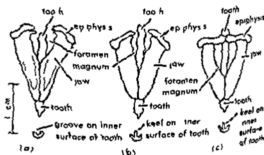


Fig 3 Lantern structure For each main type a single interradial jaw is illustrated and below each is shown a cross section of the tooth (a) *Aulodont* (b) Stirodont (c) Camarodont

and the jaw may be partly or completely roofed over by epiphyses As seen in the illustration three main types of dentition are distinguished These are (1) aulodont in which the teeth are grooved and epiphyses do not meet so that there is an open foramen magnum in the jaw, (2) stirodont in which the teeth are keeled within and the foramen magnum is open, and (3) camarodont in which the teeth are keeled and the foramen magnum is closed by the epiphyses These characters are useful in taxonomy However it has been shown by T Mortensen (1928–1951) that parallel dentitional evolution has occurred in various groups and accordingly J Durham and R Melville (1957) abandon as invalid the three orders formerly based on dentition (*see* AULODONTA CAMARODONTA STIRODONTA) As concise morphological descriptive terms, aulodont stirodont and camarodont remain valuable, and they are used in that sense in the taxonomic diagnoses which have been employed in this encyclopedia

Water vascular system This system is highly developed and greatly influences the form of the ambulacral plates It is here considered in relation to the ambs The ring vessel rests upon the lantern and bears small polian vesicles The stone canal passes upward through the coelom to the madreporite The radial vessels and their ampullae lie on the inner surface of the ambs within the test The tube feet alone emerge to the exterior by way of pores

Each tube foot traverses the test wall by means of two pores termed a pore pair One pore serves for the outward flow of hydrocoele fluid the other for inward flow The pore pairs lie on the amb plates (Fig 4) In young stages and in the adults of Cidaroida and the exocyclic orders each amb plate bears only one pore pair Such amb plates are termed simple In the older stages of other echinoids the amb plates tend to fuse into compound plates which therefore carry more than one pore pair The most common arrangement is that in which an arc of three pore pairs occurs on a plate Such a plate is termed trigeminate or oligoporous Plates with four or more pore pairs are termed polyporous One component of a plate is usually larger than the others and is termed the primary The position of the primary and the arrangement of the pore pairs provide characters used in taxonomy

Figure 4 shows types of ambulacral plates which may be defined as follows Diademoid plates may be trigeminate or polyporous with the primary immediately over the lowest element Arbacioid plates are similar but the pore pairs lie in a vertical series and the secondary elements (demiplates) are rectangular Echinoid plates differ from diademoid plates in having the primary as the lowest element The ambulacra in some exocyclic echinoids tend to change from simple meridians into petal shaped areas around the upper pole and around the mouth Such ambs are described as petaloid and the parts which surround the apical

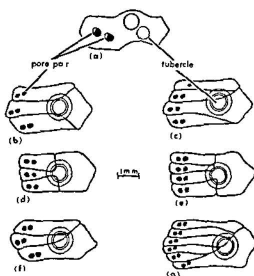


Fig 4 Ambulacral plates as used in diagnoses of the orders (a) Simple (b) Trigeminate diademoid (c) Polyporous diademoid (d) Trigeminate arbacioid (e) Polyporous arbacioid (f) Trigeminate echinoid (g) Polyporous echinoid

system are called the petals whereas those around the mouth are termed phyllodes A further development of this process occurs in the Cassiduloida where a flowerlike floscelle results

The surface of the test plates may be smooth or sculptured that is a raised pattern of ridges (epistroma) or of ridges and grooves ramifies among the tubercles *See* ARBACIOIDA, TEMNOPLEUROIDA

Some echinoids have special respiratory organs or gills attached to the peristome The gills if present usually notch the margin of the peristome The notches termed branchial slits or gill cuts may vary in size and shape or may be wanting They provide characters used in taxonomy Many exocyclic forms use the tube feet as respiratory organs

Small grasping organs pedicellariae, are well developed in echinoids in which they take the form

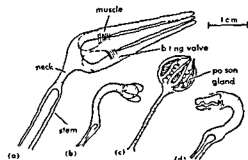


Fig 5 Pedicellariae of echinoids (a) Tridentate type (b) Triphyllous or trifoliate (c) Globiferous or gemmiform (d) Ophicephalous

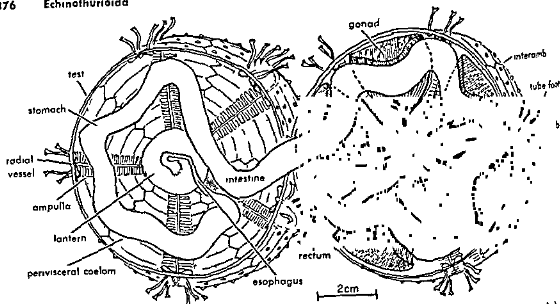


Fig 6 Dissection of echinoid by horizontal cut across the ambitus the adoral hemisphere on the left the gonads have been removed from the adoral hemisphere

of a beak carried on a stalk The beak is made up of three (sometimes only two) movable jaws operated by muscles and sometimes provided with venom glands They respond to tactile stimuli and seize any small organisms or particles which may touch the skin The intrusive material is passed from one to the other until one of the ambital pedicellariae drops it over the side The chief types are illustrated in Fig 5

Nervous system. The nervous system follows the same pattern as the water vascular system The tube feet evidently serve as tactile and taste organs A few sea urchins have photosensitive eyespots on the upper surface of the test, scattered in the ectoderm The ocular pores as noted are not sensitive to light Minute spherical stalked bodies attached to the skin in some echinoids are believed to be organs of balance

Alimentary system The alimentary canal is tubular It lies in the coelom attached to the wall of the test by mesenteries The stomach runs from the esophagus in a counterclockwise coil (as viewed from above), and the intestine retraces the route in reverse The rectum passes upward to the anus

Reproduction. Like starfishes, sea urchins seem to be sexually mature after 1 year, but continue to grow for several years The life span is unknown but may average 5–6 years in medium sized species The sexes are normally separate, although fertile hermaphrodites are known The gonads are interradial and each opens at the corresponding gonopore In exocyclic forms the gonads may be reduced in number to only three or two If a pelagic larva is present it is an echinopluteus (*see* ECHINODERMATA) In some species the young are brooded among the radioles, or in sunken petals, as in *Abatus* Spines and some other organs are regenerated after injury Autotomy and autoevisceration are unknown [H B F]

Bibliography J W Durham and R V Melville A classification of echinoids *J Paleont.*, 31(1) 242–272, 1957, T Mortensen, *Monograph of the Echinoidea*, 5 vols , 1928–1951

Echinothurioida

An order of Diadematacea with solid or hollow primary radioles, diademoid ambulacral plates noncrenulate tubercles, and the anus within the apical system (*see* ECHINOIDEA) The extant members of the two included families Pedinidae and Echinothuriidae, are all deep water forms The Pedinidae have solid radioles and simple diademoid plates They arose in the Late Triassic probably from cidarids Their fossils are abundant from the Jurassic onward, but the only surviving genus is *Caenopedina*, a brightly colored echinoid with banded radioles Pedinids probably gave rise to the Pygasteroida The Echinothuriidae have a large flexible test which collapses into a disk at atmospheric pressure, and the middle element of the diademoid plates is much larger than the other two elements Some species carry venomous spines Echinothuriids range from the Late Jurassic to present day *See* CIDAROIDA; DIADEMATACEA; PYGASTEROIDA [H B F]

Echiuroidea

A small group of wormlike animals once linked with the Sipunculoidea and Priapuloidea under the term "Gephyrea," but now regarded as a separate phylum of the animal kingdom with affinities to the annelid worms They are mainly inhabitants of the tropical and subtropical waters, living buried in the sand and mud of the sea floor from the intertidal area to the ocean depths

Their classification has always presented diff culties In the latest classification by W Fi her three orders—Echiuroinea *Xenopneusta and* Heteromyota—are recognized with two problem atic species attached The Fchiuroinea contain the families Bonellidae and Echiuridae The classifi cation of the Bonellidae is most uncertain and now includes 16 genera with only 27 species The Echiuridae are more stable with only 7 genera l ut 70 species Xenopneusta has four species and Het eromyota one Further collecting still brings to light new species often causing a modification of the classification

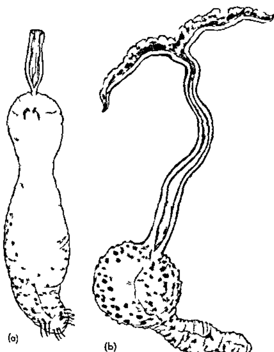
The body is saclike or sausage shaped and often highly colored Anteriorly there is a proboscis which fragments easily This may be very long and cleft at the tip as in the Bonellidae or short and flaplike in the Echiuridae It is capable of very con s derable retraction The muscles of the body wall may be an entire sheet or gathered into a varying number of bundles There is usually a pair of setae situated ventrally a short distance below the mouth and in the genera *Echiurus* and *Urechis* there are one or more rows of posterior setae

The mouth leads into the gut which is divided into *several distinct regions At the posterior end* of the body are two very characteristic structures the anal vesicles which are tubular or branched extend for varying distances into the body cavity and have a respiratory and excretory function

The sexes are separate and similar in the Echi *uridae* In the Bonellidae they are separate and very dissimilar the male being minute and para sitic within or on the female *See* ANIMAL KINGDOM

[A C S]

Bibliography F Baltzer *Echiurida* in W Ku kenthal and T Krumbach *Handbuch der Zoolo gie* vol 2 1931 W K Fisher Echiuroid worms of the North Pacific *Proc U.S Natl Museum* 96 215 292 1946 E Wesenberg Lund *Sipunculoidea* and *Echiuroidea* from West Africa *Inst roy sci nat Belg Bull* 33(42) 1957



Ech uro dea (a) *Ech urus* 1/ s ze (b) *Bonell a* 1/ s ze

Echo

A sound wave which has been reflected or otherwise returned with sufficient magnitude and time delay to be perceived in some manner as a sound wave distinct from that directly transmitted

Multiple echo describes a succession of sepa rately indistinguishable echoes arising from a single source When the reflected waves occur in rapid succession the phenomenon is often termed a flutter echo

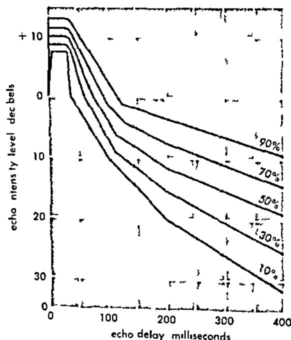
Echoes and flutter echoes are generally detri mental to the quality of the acoustics of rooms They may be minimized through the proper selec tion of room dimensions room shape and distribu tion f sound absorbing materials Flutter echoes *for example may be minimized by making the* opposite walls of a room nonparallel or by making one of the walls highly sound absorptive For a more complete discussion of the effect of room

reflected sound wave is in excess of approximately 50 milliseconds In a large auditorium where the reverberation time is of the order of seconds many reflected waves will be present (*see* REVERBERA TION) These reflections will not be troublesome if their intensity is sufficiently below that of the initial sound A relation between the approximate per cent of listeners detecting some alteration in

figure

Echoes have been put to a variety of uses in measurement problems For example the distance between two points can be measured by timing the duration required for a direct sound originating at one location to strike an object at the other point and to return an echo to the location of the initial source For the application of this principle to the detection of submarines and other submerged objects *see* SONAR

Ultrasonic echo techniques have achieved con siderable success in nondestructive testing of ma terials When an ultrasonic wave is propagated through a metal the presence of a crack or other flaw will cause a sound wave or echo to be re flected back to the initial source location Ob serving the time delay between the original sound and the perception of the echo permits the location of the flaw to be determined This technique has

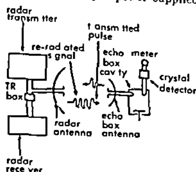


Estimated per cent of listeners disturbed by echo related to intensity of echo (relative to initial sound intensity) and the time delay between initial sound and perception of echo (*After R H Bolt and P E Doak J Acoust Soc Am 22 507–509 1950*)

been found particularly useful in such problems as examining metal castings for internal defects and determining the location of cracks in pipes or welded structures It has also been employed to locate brain tumors in man For further discussion of the application of echo reflection techniques *see* ULTRASONICS *see also* REFLECTION (SOUND) SOUND [W J C]

Echo box

A device used to check the output power and spectrum of a radar transmitter It consists of a low loss tunable resonant cavity connected to the antenna feed line through a fixed coupling circuit so that the fraction of output power supplied to the



Echo box is used in the field to check the over-all performance of a complete radar system (*From Keith Henney ed Radio Engineering Handbook 5th ed McGraw Hill 1959*)

cavity is always constant The signal level within the cavity depends on the strength of the portion of the transmitter output spectrum lying within the cavity's narrow pass band which can be tuned to traverse the entire frequency range of interest A microammeter connected through a crystal rectifier to a loop within the cavity permits reading the signal level The spectrum can be measured as the cavity is tuned to different frequencies

A single test of the performance of the entire radar system excluding only the antenna and antenna feed can be made with the cavity tuned to the carrier frequency Each transmitter output pulse causes a slowly damped oscillation in the cavity which feeds a signal back through the coupling circuit to the receiver This signal appears as an echo at the receiver output (hence the name echo box) The time required for the echo to decay to the level of the receiver noise is proportional to the logarithmic difference or difference in decibel between the transmitter power level and the receiver noise level This is an excellent over all figure of merit for the transmitter and receiver performance If a dummy load is substituted in place of the antenna to absorb the output power the radar can be tested without actually radiating which may be useful in some military situations *See* RADAR [R J B]

Echo sounder

A marine instrument used primarily for determining the depth of water by means of an acoustic echo A pulse of sound sent from the ship is reflected from the sea bottom back to the ship the interval of time between transmission and reception being proportional to the depth of the water

Echo sounders sometimes called Fathometers are used by vessels for navigational purposes not only to avoid shoal water but as an aid in fixing position when a good bathymetric chart of the area is available Some instruments are sensitive enough to detect schools of fish or scattering layers of minute marine life and are often used by commercial fishermen or marine biologists for this purpose Oceanographic survey ships use echo sounders for charting the ocean bottom Figure 1 shows an echo sounder record obtained by oceanographers of a seamount (undersea mountain) *See* SCATTERING LAYER

An echo sounder is really a type of active sonar (*see* SONAR) It consists of a transducer located near the keel of the ship which serves (in most models) as both the transmitter and receiver of the acoustic signal the necessary oscillator receiver and amplifier which generate and receive the electrical impulses to and from the transducer and a recorder or other indicator which is calibrated in terms of the depth of water An echo sounder actually measures time differences so some average velocity of sound must be assumed in order to determine the depth The frequency generally employed is in the low ultrasonic range (20 000–30

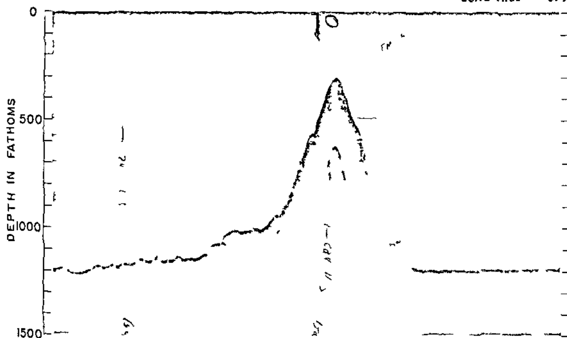


Fig 1 Echo-sounder record of a seamount in the deep ocean (Woods Hole Oceanographic Institution)

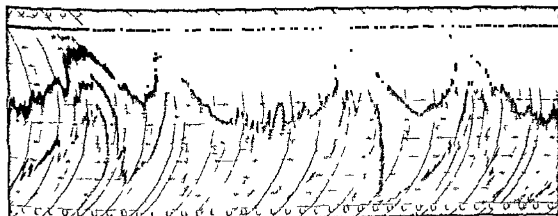
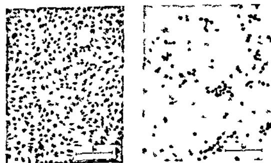


Fig 2 Typical record of bottom profile obtained by an echo sounder (USCGS)

000 cps) The depth display may be given by a trace type recorder which supplies a continuous permanent record (Fig 2) or as is the case with less expensive commercial instruments it may be a dial type display giving the instantaneous depth of water See SUBMARINE TOPOGRAPHY UNDERWATER SOUND [R W MO]

ECHO virus

Also known as enteric cytopathogenic human orphan virus They constitute a group of the enterovirus family Twenty four antigenic types exist Only certain types have been associated with human illnesses particularly with aseptic meningitis and febrile illnesses with or without rash Their epidemiology is similar to that of other enteroviruses See ANTIGEN EPIDEMIOLOGY MENINGITIS



(a) Normal monkey kidney tissue culture (b) Monkey kidney culture 3 days after infection with an enterovirus ECHO type 1 (Photomicrograph J L Melnick Baylor University College of Medicine)

ECHO viruses resemble polio viruses and Coxsackie viruses in size (about 28 millimicrons) and in many other properties They are nonpathogenic for newborn mice rabbits or monkeys but cytopathogenic for monkey kidney and other tissue cultures

Diagnosis is made by isolation and typing of the viruses in tissue culture Antibodies form during convalescence *See* ANIMAL VIRUS [J L M]

Bibliography T M Rivers and F L Horsfall Jr (eds) *Viral and Rickettsial Infections of Man* 3d ed 1959

Eclipse, astronomical

The total or partial obscuration of a celestial body by the shadow of another The Sun is eclipsed from terrestrial viewers when the Moon passes between the Sun and Earth During an eclipse of the Moon Earth passes between the Moon and the Sun and casts its shadow on the Moon The shadows cast by Earth the Moon and all other members of the solar system that receive their illumination from the Sun have the shape of a cone and extend in a direction opposite that of the Sun The axis of the shadow is the line joining the centers of the Sun and the planet or satellite considered The exterior tangents to the two bodies form the cone of umbra whereas the interior tangents form the cone of penumbra *See* SHADOW

This article deals with eclipses of the Sun Moon and Jupiter s satellites For information on related phenomena which are sometimes incorrectly defined as types of eclipses *see* BINARY STARS OCCULTATION TRANSIT (ASTRONOMY)

Solar eclipses Eclipses of the Sun are seen when the shadow of the Moon falls on Earth (Fig 1) Strictly speaking they should be called eclipses of Earth and occasionally have been so designated in the past A solar eclipse can occur only when the Sun Moon and Earth are in a nearly straight line This condition is fulfilled when the Moon is in conjunction with the Sun (at new moon) while it is near one of the nodes of its orbit The Sun takes 346 62 days to return to the same node of the Moon s orbit This period which includes one passage at each of the two nodes is called the eclipse year At least one and no more than two solar eclipses occur at each node passage Because 19 eclipse years (6585 78 days) are very nearly equal to 223 synodic lunar months (6585 32 days) the Sun Earth and Moon return to almost identical relative positions every 18 years $11\frac{1}{3}$ days ($10\frac{1}{3}$ days when there are 5 leap years in the interval) and eclipses of the Sun and Moon recur after that period This cycle is called the saros Corresponding eclipses in successive saros cycles form a series over a period averaging 1280 years Consecutive eclipses in a series have the same general characteristics There is a shift of 2–3° in latitude either north or south on Earth and a shift in longitude of approximately 120° westward at each recurrence

Types of eclipses The distance from the Moon to the center of Earth is variable When it is closest to Earth the Moon's apparent diameter exceeds that of the Sun by 2 38 ′, whereas it is less than that of the Sun by 2′40″ when the Moon is farthest away When the umbra of the Moon falls on Earth at a time when the diameter of the Moon exceeds that of the Sun, a total eclipse takes place If however, the umbra falls on Earth when the diameter of the Moon is smaller than the Sun s the Moon's disk cannot cover the Sun completely A portion of the Sun's surface remains visible as a ring around the Moon and the eclipse is called annular or ring shaped In the latter case the vertex of the cone of umbra is located outside of Earth and only the extension of the cone beyond the vertex falls on Earth

Earth moves in relation to the shadow during an f the cone t e

face is called the central line

If the vertex of the umbra lies within a few miles of Earth's surface the eclipse may be annular at the beginning of its path total near the middle and annular again at the end If only the cone of penumbra touches Earth the eclipse is partial

At any point within the path four contacts are observed namely, the first point of tangency as the Moon begins to encroach on the Sun's surface the beginning and end of total or annular phase and the end of the eclipse

Computation of eclipses Solar eclipses are computed by the method of F W Bessel The principle of the method is based on the notion of the fundamental plane This is by definition the geo

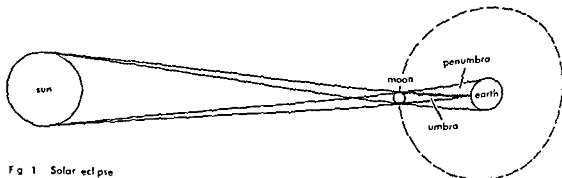


Fig 1 Solar eclipse



Solar ecl pse mult ple exposure photograph taken near St Paul M nneso a on morn ng of
30 1954 (M T nklenberg)

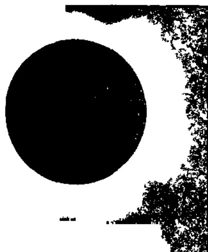


Fig 2 Solar corona photographed during the total eclipse of June 8 1937 at Canton Island Pacific Ocean (Official U S Navy Photograph)

centric plane perpendicular to the axis of shadow The position in space of the fundamental plane varies during the eclipse An instantaneous representation of the phenomenon is obtained by projection onto the plane Thus the eclipse may be computed as a function of time The Besselian elements published in the astronomical ephemerides give for selected times during each eclipse the coordinates of the axis of shadow with respect to the fundamental plane and the radii of umbra and penumbra in that plane From these elements the local circumstances of the eclipse may be derived for any place on Earth s surface

Information obtained Observations of solar eclipses provide a wealth of information The accurate times of second and third contacts and to a lesser extent the time of maximum eclipse are used to study the motion of the Moon and the irregularities of Earth's rotation The times of contacts at widely separated stations along the path of central phase when compared to the speed of shadow between those points permit the precise determination of distances over long arcs on the surface of Earth This geodetic application is impaired by the fact that solar eclipses seldom recur at the same site within a suitable time span

During the time that the extremely brilliant light of the Sun s photosphere is screened from view by the Moon it is possible to observe the fainter objects in the Sun s immediate vicinity

A quantitative check on Albert Einstein's general theory of relativity is obtained by photographing the star field in the neighborhood of the eclipsed Sun The telescope used in these observations is given suitable protection and is left at the site It is used again 6 months later to obtain a set of comparison photographs of the same star field after the Sun has moved to another region of the sky The positions of the stars on the two sets of photographs are then compared to measure the relativistic deflection of light in the vicinity of the Sun The observed shift is of the same order as Einstein s theoretical value but the lack of a good scale determination has heretofore introduced an uncertainty of about 10% in the observed value

Observations of limb darkening by means of filters or spectrograms just before second contact and just after third contact serve to evaluate the theoretical models proposed for the upper photosphere of the Sun

The best observations of the chromosphere are obtained during the few seconds near second and third contacts The gradual advance of the Moon from the visual portion of the Sun to the lower chromosphere reveals the change from the Fraunhofer (absorption) spectrum to the flash (emission) spectrum Frequent photographs taken during the few critical seconds by a prismatic camera or from a slit spectrograph with its slit tangent to the Moon s limb yield data on absolute intensities of spectral lines density gradients self absorption effects atomic abundances electron densities excitation temperatures and other related information about the chromosphere

Observational data on prominences and on the inner corona (Fig 2) are similar to those obtained by the coronagraph but the data obtained during eclipses often provide a higher degree of accuracy *See* CORONAGRAPH

Because of its extreme faintness the outer corona is most difficult to observe It is generally necessary to use an occulting disk in front of the camera to screen off the glare of the inner corona The same procedure is used for observations of the zodiacal light which is thought to be an extension of the outer corona (*see* ZODIACAL LIGHT) Another method consists of observing the outer corona and the zodiacal light while the eclipsed Sun is still below the horizon In this case Earth itself serves as a shield Polarization measurements permit observers to distinguish the nearly unpolarized F corona (light diffracted by interplanetary particles) from the polarized K corona (light scattered by electrons) This method is less reliable at the greater elongations where the type of polarization of the F corona is not known Spectroscopic observations of the intensity of absorption lines may be substituted for the polarization method inasmuch as the spectrum of the F corona shows absorption lines whereas that of the K corona does not

As the eclipse progresses the Moon gradually cuts off the ultraviolet radiation from the Sun The ensuing effects on the ionosphere of Earth are observed with suitable recording instruments The ionosphere reacts slowly making it difficult to derive from these measurements the distribution of the radiation on the Sun s disk On the other hand the observations provide a direct measure of the time that each of the ionospheric layers requires to return to equilibrium

Solar eclipses have been used for localizing discrete radio sources on the surface of the Sun and for obtaining estimates of the distribution of radio

emission from the quiet (as distinguished from the active) Sun However the resolving power of radio telescopes and interferometers has been im proved to the point where observations made when eclipses are not taking place give results of com parable accuracy *See* SUN

Lunar eclipses Eclipses of the Moon occur when the shadow of Earth falls on the Moon (Fig 3) This requires that the Sun Earth and Moon be in a nearly straight line As is the case for eclipses of the Sun the Moon must be near one of the nodes of its orbit In this case however the Moon must be at opposition that is at full moon If the Moon is immersed entirely within the cone of um bra the eclipse is total If only part of its surface is covered by the umbra the eclipse is partial If the Moon penetrates only the cone of penumbra the phenomenon is called a penumbral eclipse or lunar appulse

All lunar eclipses are visible from every point on the surface of Earth where the Moon is above the horizon at the time of the phenomenon Six contacts take place in succession during a total eclipse beginning of penumbral eclipse partial phase total phase end of total phase partial phase and penumbral eclipse The first and last contacts with the penumbra cannot be timed accurately Only a gradual decrease in the brightness of the Moon occurs during the penumbral phase and during all lunar appulses This decrease in brightness may be measured with a photometer

During the total phase the Moon does not be come entirely dark The atmosphere of Earth re fracts some of the Sun s rays into the cone of shadow thus providing the Moon with a small amount of illumination This accounts for the deep reddish color of the Moon during total phase

Because of the presence of Earth s atmosphere the edge of the shadow lacks definition (Fig 4) This interferes with the accurate timing of the con tacts and of the progress of the shadow across the features of the lunar surface The circular shape of Earth s shadow is a visible proof of Earth s rotun dity

Measurements of the brightness of the eclipsed Moon yield information on the transmission of light at various altitudes in Earth s atmosphere If an occultation (passage of a star behind the Moon) occurs during the total phase it is possible to observe both the immersion and emersion of the same star at the dark limb This affords a reliable mean value of the diameter of the Moon for use in the reduction of data from other lunar occultations Measurements of the drop in temperature on the lunar surface caused by the passage inside the shadow indicate that the material at the surface is a very poor heat conductor poorer than solid rock and comparable to crushed lava or volcanic ash *See* MOON

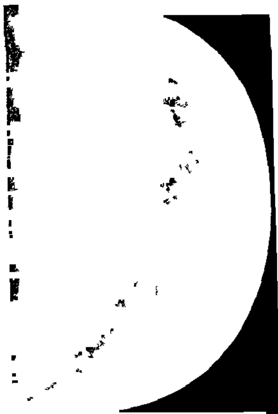


Fig 4 Partial phase of the lunar eclipse of January 29 1953 photographed at Washington D C (Official U S Navy Photograph)

Eclipses of Jupiter's moons Eclipses of the four Galilean satellites of Jupiter occur when the satel lites enter the cone of shadow of the planet When Jupiter is in or near opposition with the Sun its shadow extends in the line of sight behind the planet At that time the eclipses of the satellites cannot be seen from Earth and both the immersion

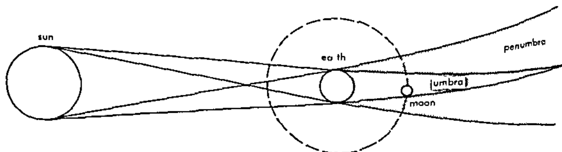


Fig 3 Lunar eclipse

and emersion of their occultations (passages behind Jupiter) are seen When Jupiter is in or near quadrature its shadow extends far to the east or west of the planet and for satellites II III and IV the eclipse and the occultation are two distinct phenomena (For an explanation of opposition and quadrature *see* PLANET) At intermediate points between opposition and quadrature the eclipse and occultation overlap with the result that only the beginning of one and end of the other are observable

Observations of the eclipses of Jupiter s satellites led Ole Roemer in 1675 to the discovery of the finite velocity of light He made this discovery from a study of the discrepancies between the predicted and observed times of these phenomena

As is the case for the Sun and Moon eclipses of the satellites of Jupiter progress gradually from a partial to a total phase The observed lapses of time between first and second contacts and between third and fourth contacts are used in conjunction with the known rate of motion of the satellites to estimate their diameters *See* JUPITER

[S D G]

Bibliography W J G Beynon and G M Brown (eds), *Solar Eclipses and the Ionosphere* 1956 F W Dyson and R v d R Woolley *Eclipses of the Sun and Moon* 1937 G P Kuiper (ed) *The Solar System* vol 1 1953 S A Mitchell *Eclipses of the Sun* 5th ed 1951 H N Russell R S Dugan and J Q Stewart *Astronomy* vol 1 1945 *Trans Intern Astron Union* 9 181-200 1957

Ecliptic

The path in the sky traced by the Sun in its apparent annual journey as Earth revolves around it The ecliptic is a great circle on the celestial sphere inclined about 23 5° to the celestial equator the angle of inclination being called the obliquity of the ecliptic *See* COORDINATE SYSTEMS ASTRONOMICAL

[G M C]

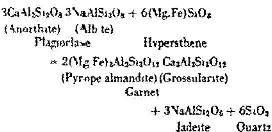
Eclogite

A class of metamorphic rocks Eclogites are different from all other rocks essentially consisting of the typical eclogite minerals a strongly green pyroxene (omphacite) and a brilliant red brown garnet (pyrope) Plagioclase feldspar has never been observed associated with the typical eclogite minerals and the eclogite minerals are not known from any other rocks but are critical for eclogites Other stable minerals occurring in eclogites are diopside enstatite olivine kyanite rutile and rarely diamond Calcite is also stable *See* METAMORPHIC ROCKS PYROXENE

Hornblende occurs in many eclogites as a hysterogenic product it is often alkali bearing Another characteristic hysterogenic product is kelyphite a peripheral alteration of garnet into pyroxene or amphibole Complete gradation is often traced from unaltered eclogite through eclogite amphibolites containing relics of garnet and omphacite together with newly generated plagioclase and hornblende to amphibolites of normal composition

Properties The properties of some of the eclogite minerals are worthy of notice Omphacite is defined as a pyroxene mixed crystal of diopside $Ca(Mg,Fe)Si_2O_6$ and jadeite $NaAlSi_2O_6$ The presence of calcium and sodium explains the absence of feldspar in eclogites (the rule of the paucity of mineral phases) The eclogite garnets are also unusual in that almandite forms extensive series of solid solutions with pyrope (up to 70 mole % that is more than in any garnet of other rocks) and with grossularite Thus the garnets exhibit an exceptionally large variation in the Ca Mg Fe ratio and a high magnesia content is typical

The following relation illustrates the difference in mineral contents between a gabbro and an eclogite of the same bulk chemical composition



$3CaAl_2Si_2O_8\ 3NaAlSi_3O_8 + 6(Mg,Fe)SiO_3$
(Anorthite) (Albite)
Plagioclase Hypersthene

$= 2(Mg,Fe)_3Al_2Si_3O_{12}\ Ca_3Al_2Si_3O_{12}$
(Pyrope almandite) (Grossularite)
Garnet

$+ 3NaAlSi_2O_6 + 6SiO_2$
Jadeite Quartz

The first is the mineral content of the gabbro the second that of the eclogite Not included is diopside which occurs in either in gabbro as a diopside augite (mixed crystal with hypersthene) in eclogite as omphacite (mixed crystal with jadeite)

The density of an average gabbro is about 3 0 that of eclogite of the same bulk composition is about 3 5 Its volume is therefore approximately 15% less than that of the gabbro The volume of jadeite is about 22% less than that of a corresponding mixture of albite and nepheline The volume of the eclogite garnet is much less than the equivalent mixtures of either pyroxene and plagioclase or hornblende and plagioclase The minor mineral constituents of eclogite—rutile kyanite and the rare diamond—are also exceptionally heavy minerals

The high density of the eclogite minerals obviously suggests that they were formed under fairly high pressure because the stability of minerals of small molar volume is favored by high pressure (Clausius Clapeyron equation)

Occurrence Eclogites are rare rocks which characteristically occur in small masses and in situations which indicate that they may have been transported from the place of origin to a foreign metamorphic environment They are found in the following forms (1) as fragments in kimberlite of the diamond pipes of South Africa and elsewhere and as inclusions in basalts associated with olivine nodules that probably derive from the subjacent sima layer of the deeper crust of the earth (2) as dikes or lenses in peridotite and serpentinite that

may be of magmatic origin (3) as blocks together with blocks of glaucophane schists caught up in serpentine (California) and (4) as localized bands in gneiss migmatite (Norway), and even in sedimentogenic mica schists and amphibolites of the Alps that are probably formed in place at loci of intense deformation at high temperature but not at great depth

The chemical composition of eclogites is re stricted to that of gabbroic and ultrabasic rocks It is possible that only rocks of high melting tempera ture remain wholly crystalline under the conditions of the eclogite facies

Eclogite may be regarded as the high pressure modification of gabbro or amphibolite P Eskola has pointed to the probable existence of a continu ous eclogite shell under the sial crust of the earth [T F W B]

Ecologic interactions

Relations between species which live together in the same community The interspecific interaction refers to the effect that an individual of one species may exert upon an individual of another species and may be physiological such as excretion of tox ins or behavioral such as fighting These interac tions may be harmful neutral or beneficial to the individual and can be described in a spectrum from positive or attraction to negative or repulsion

Symbiosis Although the term symbiosis is often restricted to an interaction that is beneficial to both species the original meaning refers to living to gether and thus includes interactions that are posi tive or negative This broad meaning has the ap proval of the American Society of Parasitologists The usage of "beneficial" and "harmful" presents various problems principally the fact

tern

g ished

An aggregation is a temporary group of animals that are attracted to some area for a specific pur pose An aggregation is an important means for utilizing the resources of the habitat since the ac tivity and call notes permit individuals to learn about the location of environmental necessities. Some aggregations may utilize physical aspects as, for example snakes sunning on a rock or may uti lize food as birds feeding on fruit trees cases spec

other

quirer

circling of vultures attracts others from many miles

Commensalism Commensal interaction refers to the joint utilization of food Literally the term means "at the same table" but the relationship is rarely equal Generally one member provides the food and the other consumes some part of it The relationship is not harmful to either party at least directly Sometimes the term is extended to include other habitat necessities such as shelter or trans portation

A spectacular example of commensalism occurs in the association of certain fish with sea anemones. These little fish *Amphiprion percula* are able to enter the anemone (*Discosoma*) and avoid harm from the poison tentacles Subsequently, the fish eat the debris of other fish that have been captured by the anemone Another example is the relation of commensal rats (*Rattus*) to humans These mam mals derive their food and shelter directly and, al though they do not normally eat "at the same ta ble," they use man's food in urban areas However they can survive very well in some wild habitats, without human assistance The rats rarely harm humans directly but may soil food and more im portant, serve as a reservoir for diseases

In many cases the relationship is facultative or unessential for survival of one member but often it is obligate or essential For example certain oli gochete worms the Branchiobdellidae attach to crayfish in an almost ectoparasitic manner and sub sist on refuse from the crayfishes' meals Obligate commensalism grades imperceptibly into parasitism and indeed, the term host is used for one partner Usually the guest lives on the outside of the host but it may live in the respiratory or digestive tract

The commensals may be about the same size or may differ greatly As the disparity in size develops there is a trend toward parasitism by the smaller This relation occurs whether the guest obtains food or shelter or transport from its host Naturally a very frequent relation is the use of the host simply as a place to live The chain of commensals may become quite long as in the case of the barnacles that perch on other barnacles that are attached to whales

Commensalism does not suggest a taxonomic re lationship Indeed plants commonly are commensal with animals, and closely related species rarely are commensal Presumably through the centuries the competition among closely related animals forced a taxonomic separation and precluded a commensal relation

Frequently, the commensals develop anatomical specialization or behavioral peculiarities in relation to their host Means of attachment processes for collection of food and senses for detecting the host are some characters that may evolve These pe culiarities may develop so elaborately that the tax onomic relations of the species are obscured or the species is utterly incapable of living without its host

Mutualism The term mutualism is now replac ing the term symbiosis to refer to relations that are beneficial to both species Such relations are vitally important to animals because the survival of the in dividual or of the species, may depend upon the success of a particular mutual interaction

Some of the mutual interactions are spectacular and intricate whereas others may be rather simple Perhaps the least complicated relation is the simple exchange of metabolic requirements In some cases this interaction is rather remote as for example

the exchange of carbohydrate and oxygen from a plant for nitrogeneous wastes and carbon dioxide from an herbivore In other cases the relation is closer as in fungal mycorhizae commonly associated with plant roots Mycorhizae have specialized processes that enter the root structure In a few cases the interaction is intimate as for example the similar exchange in a lichen which has formalized the interaction into a single structure that appears to be an individual

A somewhat different situation occurs in the interaction of certain animals and their domesticated animals and plants Some beetles ants and termites grow a number of fungi and man grows animals and plants that can no longer survive unaided in nature Although individually the domesticated animal or plant is harmed the species is maintained and thus a mutual relation occurs In many cases the relation is very specific one kind of beetle grows only one kind of fungus whereas another species grows another kind The transmission of the fungus to subsequent generations may require elaborate mechanisms

The mutual dependence of two species is extended more elaborately in the examples of digestion of an animal s food by the bacteria or protozoa in the gut For example cockroaches cannot exist without their protozoa who do the work of breaking down cellulose so that the cockroach can utilize it Many such situations exist Indeed it seems likely that most species including man are at least partially dependent upon bacteria and protozoa in the gut for digestion Some insects have elaborate structures for storing or transmitting the organisms This mutual relation occurs in plants as well The nitrogen fixing bacteria on the roots of legumes is a well known case (see NITROGEN CYCLE)

Most of the cases of mutualism are examples of a common effort to provide food either directly or indirectly However another process necessary for survival of the species namely fertilization of gametes is frequently arranged by a mutual interaction Many flowers have conspicuous devices that attract birds or insects and encourage cross pollination These mutualistic adaptations occur in some species of plants that are distantly related and thus must have developed independently Some examples are simple such as arrangement of anthers so that the pollen falls on the insect as it feeds on the nectar Other examples are more complex such as the wasp that fertilizes figs The larvae of the wasp develop with n the flower and thus the plant sacrifices some of its reproductive potential for pollination of the rest to be ensured

The variation in detail of the examples of mutualism may distract from the general trend toward interdependence At one extreme closely related forms may have some minor beneficial exchange while at the other extreme completely unrelated forms may be so dependent that they cannot exist alone This trend has led to the concept of a superorganism which is a combination of two or more species into a functional whole At the present time examples of all stages in this evolutionary story exist See PHORESY

Neutralism Neutral interactions are frequent at least at certain times or stages of life history Meadowlarks and sparrows may exist together in an area with no direct interaction except possibly the mutual use of an abundant food supply Foxes may prey upon mice during the winter but ignore them during the summer in favor of grasshoppers and berries All the species which are members of an ecologic community however have some indirect interactions While the term neutral interaction is somewhat self-contradictory it is helpful to express the fact that some potential interactions result in no evident effect on either species at any time or only at certain times Furthermore the interaction can change from neutral to positive (aggregation commensal symbiotic) or to negative (avoidance competition predation parasitism) under various conditions

pulsion Small animals such as crows surrounding a carcass may learn to avoid the larger species like vultures which drive them away thus harmful results are prevented The avoidance may be direct as in the previous example or may be based on signs such as trails or feces Among higher forms especially vertebrates the reaction may be inherited The avoidance of a potentially harmful interaction obviously improves survival of the individual

Competition This interaction results when sev

before the other thereby preventing his use of the resource Direct competition occurs when an elk drives a deer away from an item of food and indirect competition occurs when the elk eats an item of food in the summer that the deer would eat in the winter It is at once apparent that competition

among individuals of the same species or interspecific that is among individuals of different species Actually in many cases the results may be similar It matters little to a deer whether another deer or

to reproduce Thus while competition rarely leads

simple means naturally is

when an individual may flee but return at a later time to obtain the desired item Animals of the same species or even different species may formalize this device into a social rank or hierarchy that regulates the order of precedence (*see* SOCIAL ANIMALS) Under this arrangement a low ranking individual simply waits his turn at the food supply This has been studied in domesticated chickens who form a definite rank in which the top ranking individuals take their share first If the supply is inadequate the low ranking individuals do without If shortage becomes more acute the low ranking individuals flee or eventually die The use of social organization to systematize the effects of competition occurs in many variations in many species but is most highly developed in vertebrates While the result may be disastrous to the individual the consequences for the population may be beneficial If there exists only enough food for two roosters it is better for survival of the population to let the two have enough

l l f d ll among

beneficial for the species

Another behavioral device territorialism may mitigate the effects of competition Individuals or pairs defend an area against intruders of the same species or of different species Although most cases of *territorial interactions* occur among individuals of the same species in some cases the interactions occur among members of various species In the winter red headed woodpeckers defend an area in which they store acorns The individual drives out other red headed woodpeckers and also titmice and *chickadees* Under these circumstances the interaction is both intra and interspecific

Examples of competitive *interactions are* surprisingly difficult to document quantitatively even in *the laboratory* The behavioral devices of flight rank *and* territorialism mitigate the effects in natural *populations* to such an extent that only gradual changes *occur Flickers* have decreased in numbers in many areas since *the European* starling arrived yet direct combat for nest holes is rarely seen and it is *difficult* to measure the interaction Laboratory experiments on *the competition* of beetles mice viruses and a few *other* forms show results that resemble the interactions *in nature* The precise effect of an interaction is *difficult* to predict even in the laboratory because it depends *upon many* variables

Predation This is the killing and eating of an individual of one *species by an individual* of another species A conventional *example is the* killing of a mouse by a cat However the term refers also to the killing of individuals by a group, such as results from the attacks of a pack of wolves upon a deer

Predation is also used to refer to a process in a population It is the statistical effect on the population by the various predators Thus predation is said to reduce a population or to alter the age composition In these cases the mass effect of some or all of the predators on the population is being considered

— are commonplace ragonfl es catch its starfi h trap dation on indiv

ture claims viduals exists primarily in terms of natural selection It has been known even before Darwin that predators may select certain types of individuals from among a population but only recently have quantitative data become available In England predators captured light colored moths at a higher rate than dark colored individuals Thus during many generations selection will favor the dark forms

The evolution of devices which improve ability to capture and kill the prey has produced some unusual structures Simple examples are claws and teeth while more complex examples are the pits of ant lions and the traps of several plants like the Venus flytrap Obviously the guns nets and traps created by humans are merely extensions of devices for predation It is of course apparent that some devices used for predation can also be used in competitive interactions

The result of the interspecies interaction of predation when individuals are concerned is relatively simple The prey is killed or escapes *and the predator* gets food or searches elsewhere However a number of compensatory processes complicate the result on the population of the prey Indeed a continuous spectrum of results on the population can be prepared from inadequate predation on the one hand to excessive predation on the other hand The terms inadequate and excessive refer to the welfare of the prey species Examples *of* inadequate predation occur in many fish populations Predators such as man or herons may be *unable to remove* enough fish *to* prevent a population increase that reduces the food supply with the result that the individual fish are stunted and may not reproduce An increase in predation may *reduce the population* numerically but result in an increase in size and in the rate of reproduction

In other cases predation has no *measurable* effect on the population other than the removal of particular individuals since predation *does not* affect the number of prey, except *momentarily* if the total mortality rate remains *constant*. Increased *hunting did not affect pheasant* population in Michigan partly because the pheasants learned to avoid hunters and partly due to a *compensatory* decrease in mortality from other causes

However in some cases predation can *be clearly* associated with a decline in population such as the reduction in the number of mosquitoes in an area by the mosquito fish *Gambusia* or the reduction of rabbits in Australia by the virus disease myxomatosis Humans of course have reduced *many species to low levels*

The extreme effect of the interaction is the extinction of a prey species by a predator Man has exterminated many vertebrates in recent years However the extinction of one vertebrate by another vertebrate other than man is rare The only examples adequately proven occur when rats or mongooses exterminate some birds on islands Disease provides some spectacular cases of extinction such as the extermination of the American chestnut by blight a virus

These examples demonstrate the spectrum of interspecific interactions that occur under the general term of predation While they may harm the individual in many cases the survival of the species is improved

Parasitism This is a variety of relations in which a small species lives in or on another species and usually derives food or shelter from it As might be expected relations that are mutual or commensal grade imperceptibly into parasitic Indeed the same two species may have mutual or commensal relations at one time and parasitic at another The distinction among the three types of symbiosis is based upon the relation to the survival of an individual or at least to its welfare If one individual is harmed the relation is called parasitic However usually only certain of the parasitized individuals are harmed or in the case of ectoparasites the harm is trivial

Two types of parasites are distinguished ectoparasites which are external and endoparasites which are internal The parasite needs to spend only a part of its life cycle on or in the host In most examples the parasite spends only a fraction of time in any one host but may inhabit two or more hosts during its life

Some general relations among parasites may be mentioned Parasites may be facultative or obligate The former are able to exist independently but may be parasitic on certain occasions whereas the latter cannot survive at certain stages without the host Parasites may be specific living only in a certain host species or nonspecific living in many hosts

Parasitic interrelations may exist between practically any group of plants and animals Plants may parasitize plants or animals and animals may parasitize animals or plants The taxonomic relations may be close in some cases or distant in others Indeed every conceivable combination occurs even the example of one sex parasitizing the other as in the angler fish

A rather aberrant type of parasitism called social parasitism occurs in birds The female lays her eggs in the nests of other species and permits the foster parents to raise the young This habit has appeared independently in five families

While the interaction of parasitism may be harmful to the individual it is not necessarily harmful to the species As is true for predation the effect of parasitism may prevent overpopulation and stimulate the evolution of new structures or adaptations A value judgment about its effect on the species is precarious *See* POPULATION DISPERSAL [D E D]

Bibliography W C Allee et al *Principles of Animal Ecology* 1949 L R Dice *Natural Communities* 1952

Ecological systems, energy in

Energy is defined as equivalent to work having the dimensions L^2M/T^2 and being measured in ergs It occurs as potential and as kinetic energy in static and dynamic systems respectively Its liberation results in its appearance as heat light mechanical work chemical change or electric current All forms of energy are potentially interconvertible and subject to the laws of thermodynamics which state that energy is neither created nor destroyed and that such conversions occur spontaneously only when the result is to produce energy in a more dispersed form than before This is equivalent to saying that the over all result of all energy changes is to increase the entropy of the system as a whole Entropy is a measure of uniformity increased probability or lack of organization

Ecological systems are highly improbable ones in which entropy production of the universe as a whole is used to create local pockets of concentrated energy—the organisms and their stored foods The accumulation of energy in this way is another aspect of the accumulation of information the techniques of information theory can be applied to the study of energy flow in ecosystems (B Patten) *See* ENTROPY INFORMATION THEORY (BIOLOGICAL APPLICATIONS)

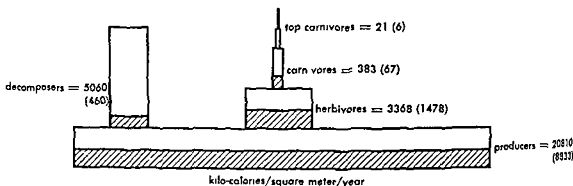
Measurement of energy flow Because energy liberation and oxidation of food are quantitatively [illegible]

type of food (carbohydrate fat or protein) is known When it is not the respiratory quotient (RQ) must be determined from any two of these quantities *See* RESPIRATION

[illegible]

insects and mammals of equal sizes respire at very different rates One result of this is that biomass is a poor indicator of the energy flow capacity of organisms

Path of energy flow Because energy enters ecological systems only through photosynthesis and because the energy is lost to the system after it has been liberated by oxidation of food energy [illegible] the quan [illegible] tion of [illegible] ow diagrams and energy pyramids (see illustration) Energy flow diagrams show the sequential fate of energy contained in *food whereas energy pyram*



Energy pyramid for S lver Springs Florida on an annual basis Portion of total energy flow which is actually fxed as organ c biomass and potentially available as food for other populations in the next trophic level indicated by fgures in parentheses and by the shaded portion of each tier About one half the total assimilation ends up in the bod es of organ sms (the rest being lost as heat or export) except for the decomposers level where the percentage loss in respiration is much greater (*After H T Odum in E P Odum Fundamentals of Ecology 2d ed Saunders 1959*)

which were introduced by E Odum demonstrate the quantitative division of metabolism between the trophic levels and the relation between biomass and energy *See* BIOLOGICAL PRODUCTIVITY FOOD CHAIN [A MC]

Bibliography S Brody *Bioenergetics and Growth* 1945 E P Odum *Fundamentals of Ecology* 2d ed 1959 B C Patten An introduction to the cybernetics of the ecosystems the tropic dynamic aspect *Ecology* 40(2) 221 1959

Ecology

A study of the relation of organisms to their environment or in more simple terms environmental biology Because ecology is concerned especially with the biology of groups of organisms and with functional processes on the lands in the oceans and in fresh waters it is more in keeping with the modern emphasis to define ecology as the study of the structure and function of nature (mankind is considered to be a part of nature)

Ecology is one of the basic divisions of biology which are concerned with principles or fundamentals common to all life Morphology physiology genetics embryology and evolution are examples of other basic divisions of the science of biology One may also divide biology into what may be called taxonomic divisions which deal with the morphology physiology and ecology of specific kinds of organisms Zoology and botany are large divisions of this type whereas mycology entomology and ornithology are divisions dealing with more limited groups of organisms Thus ecology is a basic division of biology and as such is also an integral part of any and all of the taxonomic divisions Both approaches are profitable It is often productive to restrict studies to certain taxonomic groups because different kinds of organisms require different methods of study and some groups of organisms are of greater interest to man than others It is also important to seek and to test unifying principles which may be applicable to nature as a whole *See* ANIMAL COMMUNITY, PLANT COMMUNITY

Approaches to ecology From the over all aspect, ecology may be subdivided in three ways which also represent three approaches of study taxonomic habitat, and population

Taxonomic approach As indicated in a previous paragraph the subject may be subdivided taxonomically as insect ecology or bird ecology

Habitat approach A convenient subdivision may be based on the kind of environment or habitat to be considered or studied Marine ecology fresh water ecology and terrestrial ecology are the three broad divisions from this point of view estuarine ecology stream ecology or grassland ecol gy represent more restricted interests From the standpoint of terrestrial ecology the most instructive subdivision is in terms of the concept of level of organization For convenience a biological spectrum can be visualized as follows protoplasm cells tissues organs organ systems organisms populations communities ecosystems and the biosphere Ecology is concerned largely with the levels beyond that of the individual organism

Population approach In ecology the term population originally used to denote a group of people is broadened to include groups of individuals of any one species of organism (*see* POPULATION DYNAMICS) Community in the ecological sense sometimes designated as biotic community includes all of the species populations of a given situation (*see* COMMUNITY) The community and the nonliving environment function together as an ecological system or ecosystem (*see* ECOSYSTEM) The portion of the earth in which ecosystems can operate that is, the soil air and water is conveniently designated as the biosphere (*see* BIOSPHERE) The term autecology is often used to refer to the study of environmental relations of individuals or species, whereas the term synecology refers to the study of groups of organisms such as communities

Some attributes obviously become more complex and variable during the procession from cells to ecosystems however it is an often overlooked fact that other attributes may become less complex and less variable from the small to the large unit The reason for this is that homeostatic mechanisms which may be defined as checks and balances or forces and califorces operate all along the line to produce a certain amount of integration and smaller units function within larger units (*see* HOMEOSTASIS) For example the rate of photosynthesis of a whole forest or a whole corn field may be less variable than that of the individual trees or corn plants within the communities because when one individual or species slows down another may speed up in a compensatory manner

Furthermore it is important to emphasize that findings at any one level aid in the study of another level but never completely explain the phenomena occurring at that level The old saying that the forest is more than a collection of trees will illustrate very well what is meant here For a full understanding of the forest it is necessary to study both the trees as separate units and also the forest as a whole During recent years ecology has made the most progress not as a result of intensive study at a single level but by coordinated and simultaneous attack at all levels *See* FRESH WATER ECOSYSTEM MARINE ECOSYSTEM TERRESTRIAL ECOSYSTEM

Coral reef study The importance of studying both the part and the whole may be illustrated by the following example A coral reef represents one of the most beautiful and best adapted ecosystems Corals are small animals that have tentacles adapted to seize small animals called zooplankton which are in the water Embedded in the tissues of the coral animal are numerous small plants or algae Some years ago C M Yonge carried out a carefully planned series of experiments with isolated coral colonies in tanks in an effort to clarify the relationship between corals and their contained algae He found that when the corals were supplied with abundant zooplankton they thrived and grew normally when all of the algae were killed by keeping the colonies in the dark On the basis of these experiments Yonge concluded that the algae do not contribute to the well being of the coral and therefore are of no importance in the reef building activities of corals Some years later H T Odum and E P Odum were able to measure the metabolism of an intact coral reef and thus determine the amounts of food which corals require It was soon evident that there were not enough zooplankton in the surrounding infertile oceans to account for the large population and rapid growth of the coral reef It was suggested therefore that the corals must indeed obtain some of their food from the algae Other investigators became interested in the problem and recently L Muscatine and C Hand (1958) showed by the use of carbon 14 isotopic tracer that food does indeed diffuse from the algae which manufacture food by photosynthesis to the coelenterate Thus it was proved that what was true of an isolated colony in a tank was not entirely true for the coral living in its natural ecosystem *See* CORAL REEF

Functional ecology Returning to the original definition of ecology namely the study of structure and function of nature it would be well to point out that until fairly recently the descriptive aspect was largely emphasized Ecologists often described how nature looked Now equal emphasis is being placed on what nature does because the changing face of nature can never be understood unless her metabolism is also studied Consideration of function brings the small organisms which may be inconspicuous but very active into true perspective with the large organisms which may be conspicuous but relatively inactive It is evident that as long as a purely descriptive viewpoint is maintained there is little in common between such structurally diverse organisms as trees birds and bacteria In real life however all these are intimately linked functionally in ecological systems according to well defined laws Likewise from a descriptive standpoint a forest and a pond have very little in common yet both of these environmental systems function according to exactly the same principles Thus general ecology is essentially an ecology of function

THE ECOSYSTEM

The ecosystem is the basic functional unit in ecology because it includes both the organisms and the nonliving environment each influencing the properties of the other and both necessary for maintenance of life on the earth Any area of nature that includes living organisms and nonliving (abiotic) substances interacting to produce an exchange of materials between the living and nonliving parts is an ecological system or ecosystem

Biotic components From a functional standpoint an ecosystem has two biotic components an autotrophic component able to fix light energy and manufacture food from simple inorganic substances and a heterotrophic component which utilizes rearranges and decomposes the complex materials synthesized by the autotrophs From a structural standpoint it is convenient to recognize four constituents comprising the ecosystem as shown in Fig 1 (1) abiotic substances basic elements and compounds of the environment (2) producers the autotrophic organisms largely the green plants (3) the large consumers or macroconsumers heterotrophic organisms chiefly animals which ingest other organisms or particulate organic matter and (4) the decomposers or microconsumers (also called saprobes or saprophytes) heterotrophic organisms chiefly the bacteria and fungi which break down the complex compounds of dead protoplasm absorb some of the decomposition products and release simple substances usable by the producers As shown in Fig 1 a pond

is a good example of an ecosystem which exhibits a recognizable unity both in regard to function and structure

The stratification of autotrophic and heterotrophic functions and basic trophic levels is the same for land and water systems but the kinds and relative sizes of the organisms are different On land the individual producer organisms tend to be relatively large in relation to consumers whereas in water the producers are often microscopic in size Although the biomass of the land vegetation per unit of area may be much greater than that of the biomass of aquatic phytoplankton photosynthesis of the latter can be as great if light and nutrients are equivalent Marshes ponds and shallow margins of lakes and the oceans have a trophic structure intermediate between these two systems because both large and small producers may be present

The concept of the ecosystem is and should be a broad one its main function in ecological thought being to emphasize obligatory relationships, interdependences and casual relationships Ecosystems may be conceived and studied in various sizes A *small* pond a large lake a tract of forest or even a small aquarium may provide a convenient unit of study As long as the major components are present and operate together to achieve some sort of functional stability even if for only a short time the entity may be considered an ecosystem Throughout the entire biosphere ecosystems have a very similar functional makeup, though they may differ markedly in structural features and degree of stability

Very frequently the basic functions and the organisms responsible for the basic processes in an ecosystem are partially separated in space and also in time The autotrophic and heterotrophic components are often stratified one above the other and there is often a considerable delay in the heterotrophic utilization of the products of the autotrophic organisms For example, photosynthesis predominates in the aboveground portion of a forest ecosystem Only a part, often very small of the food manufactured in the upper layers of the forest is immediately and directly used by the plants and by herbivores and parasites which feed on foliage and new wood Consumption of much of the synthesized food material in the form of leaves wood and stored food in seeds and roots is delayed until it eventually falls to the ground and becomes part of the litter and soil which constitute a well defined heterotrophic stratum A similar functional stratification and time lag may be observed in the pond ecosystem (Fig 1), the phytoplankton comprise the well defined autotrophic layer whereas the sediments constitute the heterotrophic system

The ecological categories which have been discussed above are ones of function rather than of species There are no hard and fast lines between such categories as producers, consumers and decomposers because some species of organisms oc

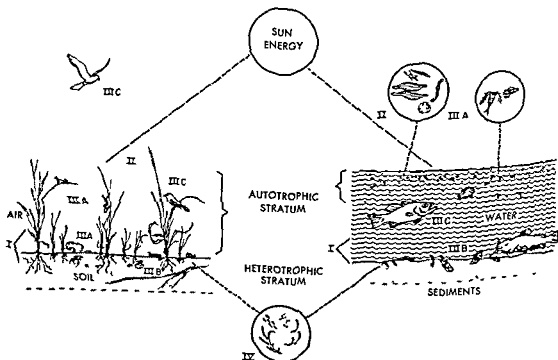


Fig 1 Comparison of the trophic structure of a simple terrestrial ecosystem (a grassland) with an open water aquatic ecosystem (either fresh water or marine) Basic units of the ecosystem are I, abiotic substances (basic inorganic and organic compounds); II, producers (vegetation on land small floating algae, or phytoplankton in water) III, macroconsumers (A = direct herbivores B = detritus eaters or saprovores C = carnivores) and IV, decomposers (bacteria and fungi of decay)

cupy intermediate positions in the series and others are able to shift their mode of nutrition according to environmental circumstances For example certain species of algae and bacteria are able to function either as autotrophs under certain conditions or at least partly as heterotrophs under other conditions The separation of heterotrophs into large consumers and small decomposers is arbitrary but is justified in practice because of the different study methods required Organisms which are usually designated as decomposers (bacteria and fungi) are relatively immobile (usually imbedded *in the medium* being decomposed) and are very small with high rates of metabolism and turnover They obtain their energy by heterotrophic absorption of decomposition products Their specialization is more evident biochemically than morphologically consequently their role in the ecosystem cannot be determined by such direct methods as looking at them or counting their numbers Rather their actual functions must be measured Organisms which are designated as consumers or macroconsumers obtain their energy by heterotrophic ingestion of particulate organic matter These are largely the animals in the broad sense In contrast to the *decomposers the* macroconsumers *are larger have* slower rates of metabolism and are more readily studied by direct means They tend to be morphologically adapted for active food seeking or food gathering with the development in higher forms of complex sensory and neuromotor as well as digestive respiratory and circulatory systems Much may be inferred about the functioning of the macroconsumer groups by observing them and counting their numbers Even in this case however it is necessary to devise means of assaying rate functions if a full understanding of their role in the ecosystem is to be obtained

It is possible *to have a sizable ecosystem* containing only producers and decomposers pioneer aquatic communities for example may be composed of algae and heterotrophic microorganisms Almost everywhere on earth however the macroconsumers or animals invade sooner or later and play a prominent role in the functioning of the ecosystem It is convenient to consider soil fallen logs or deep sea basins as ecosystems (because they show *consistent structural* and *functional characteristics*) provided it is recognized that these systems consist only of the heterotrophic components and are therefore ecologically incomplete

Biotic influence on environment It is understood the physical environment controls the activities of organisms It is not always realized that organisms also influence and control the abiotic environment in many ways Changes in the physical and chemical *nature of the inert materials of the* earth are constantly being affected by organisms which return new compounds and isotopes to the nonliving environment For example the chemical content of sea water and of air is largely determined by the action of organisms Plants growing on a sand dune build up a soil radically different from the original substrate Thus the biosphere is important not only as a place in which living organisms can exist but also as a region in which the incoming radiation *energy of the sun brings about* fundamental chemical and physical changes in the inert material of the earth chiefly through the functioning of various ecosystems

Homeostatic mechanisms The existence of homeostatic mechanisms at different levels of biological organization is well known and has already been mentioned It is important to emphasize that equilibrium between organisms and environment *may be maintained by factors* which resist change in the ecosystem as a whole Much has been written about this balance of nature but the fundamentals involved are not yet clearly understood These mechanisms include those which regulate the storage and release of nutrients as well as those which regulate the growth of organisms and the production and decomposition of organic substances Many organisms particularly decomposer and producer groups release organic substances into the environment during their growth processes These substances often have a profound influence on other organisms and on the regulation of func*tion of the whole ecosystem Some substances* may be said to be antibiotic in that they inhibit the growth of other organisms whereas other substances may be stimulatory as for example various vitamins and other growth promoting substances Such external metabolites may be considered to be ectocrines or environmental hormones in that there is growing evidence that many such substances influence and control the functioning of the ecosystem in the same general manner that endocrines control metabolic rates within individual organisms Much needs to be learned about the specific action of these substances

As a result of the evolution of the central nervous system and brain man has gradually become a most powerful organism as far as the ability to modify the ecosystem is concerned Man s power to change and control unfortunately seems to be increasing faster than his realization and understanding of the profound changes of which he is capable Al

trod iced into a stream at a moderate rate for example the system is able to purify itself and

ural homeostatic mechanisms have been involved are included the stream may be permanently altered or even destroyed as far as usefulness to man *is concerned Consequently the concept of the eco*system and the realization that mankind is and must always be part of an ecosystem and that he has increasing power to modify these systems are concepts basic to modern ecology These also are points of view of extreme importance to human affairs generally Conservation of natural resources

one of the most important practical applications of ecology must be built around these points of view Thus if understanding of ecological systems and moral responsibility among mankind can keep pace with man s power to effect changes the old practice of unlimited exploitation of resources will give way to unlimited ingenuity in perpetuating a cyclic abundance of resources *See* CONSERVATION OF RE SOURCES WATER POLLUTION *see also* ECOLOGY AP PLIED

ENERGY FLOW

Everyone is familiar with the fact that the kinds of organisms to be found in any particular part of the world depend on the local conditions of exist *ence and on the* geography because each major re gion of the earth especially if isolated from other regions tends to have its own special flora and fauna (*see* BIOGEOGRAPHY BIOTIC ISOLATION Is

[illegible] parts of the globe where the basic environment is similar *The species of grasses in the temperate semiarid part of Australia* are largely different from those of a similar climatic region in North America but *they perform the same* basic function as producers *in the ecosystem* Likewise the grazing kangaroos *of the Australian* grasslands are ecological equiva lents of the grazing bison or the cattle which have replaced them on North American grasslands The *kangaroo and bison* although not closely related taxonomically *occupy* the same ecological niche in the sense *that they have* a similar functional po sition in the ecosystem in a similar type of habitat It is also *true that* the same species may function differently *that is* they may occupy different niches *in different habitats* The point to emphasize is that a list of *species* to be *found* in an area is not sufficient information in itself to determine how the biotic community *works For a full* understand ing of nature rate *functions must* also be investi gated

Energy and materials In any ecosystem the number of organisms *and the rate* at which they live depends on the *rate at which* energy flows through the system and the rate at *which materials*

[illegible]

[illegible] many times between living and non living entities that is to *say they* may be used over and over again On the *other hand* energy is used once by a given organism or *population* and

[illegible]

[illegible] materials in the *ecosystem is of* primary concern to ecologists The one-way flow of energy and the circulation of materials *are the* two great principles or laws of general *ecology* because these principles apply equally to all *environments* and all

organisms including man *See* ECOLOGICAL SYS TEMS ENERGY IN

Energy flow A simplified energy flow diagram which might, in principle, be applied to any eco system is shown in Fig 2 The boxes represent the population mass or biomass whereas the "pipes" depict the flow of energy between the living units Only about one half of the average sunlight imping ing upon green plants (producers) is absorbed by the photosynthetic machinery and only a small por tion of absorbed energy, about 1 5% in productive vegetation is converted into food energy The to tal assimilation rate of producers in an ecosystem is designated as primary production or primary productivity Gross primary production (P_G in Fig 2) is the total amount of organic matter fixed in cluding that used up by plant respiration during the measurement period, net primary production is organic matter stored in plant tissues in excess of respiration during the period of measurement Net production represents food potentially available to heterotrophs In Fig 2 net primary production is represented [illegible] producer component [illegible]

favorable [illegible]

respiration may account for as little as [illegible] gross production However, under most conditions in nature net production is less than 90% of gross *See* BIOMASS, BIOLOGICAL PRODUCTIVITY

Energy transfer. *The transfer of* food energy from the source in plants through a series of or ganisms with repeated eating and being eaten is known as the food chain In complex natural com munities organisms whose food is obtained from plants by the same number of steps are said to be long to the same trophic level Thus green plants occupy the first trophic level that is the producer level plant eaters (herbivores) the second level or the primary consumer level, carnivores which eat [illegible]

this trophic classification is one of function and of species as such a given species population may occupy one trophic level or more according to the source of energy actually assimilated *See* FOOD CHAIN

Energy degradation At each transfer of energy from one organism to another, or from one trophic level to another a large part of the energy is de graded into heat as required by the second law of thermodynamics The shorter the food chain or the nearer the organism to the beginning of the food chain the greater will be the available food energy As shown in Fig 2 the energy flows are greatly reduced with each successive trophic level whether the total flow is considered or the produc tion *P* or respiration *R* components The red[uction] with each link in the food chain is about one order of magnitude Thus for every 100 calories of net plant production in a stable community about 10 kcal would probably be reconstituted into primary consumers the $P/P_{t\ 1}$ efficiency ratio shown in

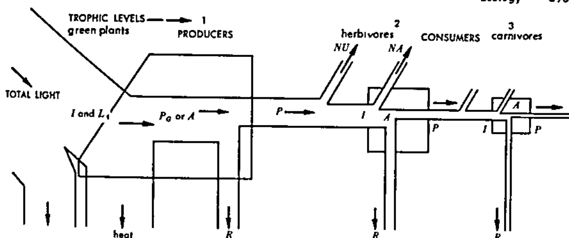


Fig 2 A simplified energy flow diagram of an ecosystem L_A absorbed light P_G total photosynthesis (gross production) P production I energy intake R respiration A assimilation NA ingested but not assimilated and NU not used by trophic level shown (After E P Odum Fundamentals of Ecology 2d ed Saunders 1959)

Fig 2, which then becomes available to carnivores which in turn might convert 1 kcal of the 10 into new organic materials Although a small amount of the primary food energy might be involved in a number of transfers it is evident that the amount of food remaining after two or three successive transfers is so small that few organisms could be supported if they had to depend entirely on food available at the end of a long food chain For all practical purposes then the food chain is limited to three or four links

Ecological efficiencies Ecological efficiencies may be expressed by formulas which indicate the ratios between and within trophic levels

$$E_I = \frac{I_T}{I_{T-1}} \quad \text{or} \quad \frac{I_2}{I_1} \qquad E_E = \frac{A_T}{A_{T-1}}$$

$$E_P = \frac{P_T}{P_{T\ 1}} \qquad E_U = \frac{I_T}{P_{T\ 1}} \quad \text{and} \quad \frac{A_T}{P_{T\ 1}}$$

Intake growth production and utilization efficiencies between trophic levels are expressed by the above equations The Lindemann efficiency or the efficiency of trophic level intake E is expressed as the ratio between two successive trophic levels the subscripts indicate the trophic levels being compared Growth efficiency E_E is the ratio between the assimilation at one trophic level with that of the preceding trophic level Trophic level production E_P is the ratio of production rates at different trophic levels The efficiency of utilization E_U is the ratio of intake at one level to the production rate at the preceding level or it may be expressed as the assimilation rate A_T to the production rate $P_{T\ 1}$

$$E_G = \frac{P_T}{A_T} \quad \text{and} \quad \frac{P_T}{I_T} \qquad E_A = \frac{A_T}{I_T}$$

Ecological efficiencies within trophic levels are measured as tissue growth and assimilation efficiencies Tissue growth efficiency E_G is the ratio between the production and assimilation rates at the same trophic level or as the production intake ratio Assimilation efficiency E_A is expressed as a ratio of the rate of assimilation to the energy intake or consumption

Energy flow models Division of heterotrophs into large and small categories that is macroconsumers and decomposers is arbitrary in terms of function but convenient in terms of analysis and study In the simplified diagram of Fig 2 bacteria and fungi which decompose plant tissues and stored plant food would be placed in the primary consumer box along with herbivorous animals likewise microorganisms decomposing animal remains would go along with the carnivores However because there is usually a considerable time lag between direct consumption of living plants and animals and the ultimate utilization of dead organic matter not to mention the metabolic differences between animals and microorganisms a more realistic energy flow model is obtained if the decomposers are placed in a separate "box connected by appropriate energy flows to the other components For example the NU (not utilized) and NA (not assimilated) flow components if not exported from the system would ultimately be utilized by decay organisms which in turn might supply at least part of the food used by such macroconsumers as detritus-feeding or filter feeding animals Organic excretions sugars amino acids and other organic materials which leak out of organisms into the environment may be considered to be a part of production because these materials are ultimately consumed by microorganisms As already discussed such metabolites may also have important functions as chemical regulators in the ecosystem

Standing crop and energy flow As has already been indicated the boxes" in Fig 2 represent the biomass of organisms functioning at the trophic level indicated The number of organisms per unit area at any one time or the average quantity over

a period of time may be conveniently designated as the standing crop Thus these boxes represent the living weight of the standing crop The relation ship between the boxes and the pipes that is between standing crops and the energy flows *P A* or *I* is of great interest and importance The energy flow must always decrease with each successive trophic level Likewise in many situations the standing crop also decreases However standing crop biomass is much influenced by the size of the individual organisms making up the trophic group in question In general the smaller the organism the greater will be the rate of metabolism per gram of weight (inverse size metabolism law) Thus 1 g of bacteria may have an energy flow equal to many grams of cow or 1 g of small algae may be equal in metabolism to many grams of tree leaves Con sequently if the producers of an ecosystem are composed largely of very small organisms and the consumers are large then the standing crop bio mass of consumers may be greater than that of the producers even though of course the energy flow of the latter must average greater assuming that food used by consumers is not being imported from another ecosystem Such a situation often exists in marine environments of moderate depths because bottom invertebrate consumers (clams crustacea and echinoderms) and fish often outweigh the phy toplankton (microscopic floating algae) on which they depend By harvesting at frequent intervals man as well as the clam may obtain as much food (net production) from mass cultures of small algae as he obtains from a grain crop which is harvested after a long interval of time However the standing crop of algae at any one time would be much less than that of a mature grain crop To summarize standing crop biomass is usually ex pressed in terms of grams of organic matter grams of carbon or kcal per unit of area or volume Pro ductivity is expressed as grams or calories per unit of time As indicated by the above examples these two quantities should not be confused the relation ship between the two depends on the kind of organ isms involved

Gross production and respiration The relation

...

tion of the ecosystem and in predicting future events One kind of ecological climax or steady state exists if the annual production of organic matter equals total consumption that is $P/R = 1$ and if exports and imports of organic matter are either nonexistent (as in a self sufficient climax) or equal In a mature tropical rainforest the bal ance may be almost a day by-day affair whereas in mature temperate forests an autotrophic regime in summer is balanced by a heterotrophic regime in winter Another type of steady state exists if gross production plus imports equals total respiration as in some types of stream ecosystems or if gross production equals respiration plus exports as in stable agriculture

Structure and species composition Seasonal fluctuations and annual shifts related to short term meteorological or other cycles in the physical en vironment occur in almost all ecosystems However the over all structure and species composition of steady state communities tend to remain the same although it is not yet certain that this is always true If production and heterotrophic utilization are not equal that is P/R is greater or less than 1 with the result that organic matter is either accu mulated or depleted the community can be ex pected to change by the process of ecological suc cession Succession may proceed either from an ex tremely autotrophic condition ($P > R$) or from the extremely heterotrophic condition ($P < R$) to wards a steady state condition in which P equals R Organic development in a new pond and the de velopment of a forest on a fallow field are examples of the first kind of succession In these situations the kinds of organisms change rapidly from year to year and organic matter accumulates. Progres sive changes in a stream polluted with a large amount of organic sewage exemplify the other type of succession in which organic matter is used up faster than it is produced in situ The two types of succession may be contrasted on a small scale in the laboratory by starting with a culture of algae on the one hand and a hay infusion on the other Because net community production (the ac cumulation of organic materials in plants and ani mals above that consumed by the community) and thereby the potential yield or harvest to man is high the first type of succession (autotrophic re gime) is much favored by man where specific prod ucts such as lumber or game are desired from natu ral ecosystems However, such situations are often not stable and tend to change to a situation with a resultant reduction in the yield to man For ex ample fishing is frequently good for a number of years in a new pond when food chains are relatively simple and a large portion of the energy flow goes into rapidly growing fish As the pond becomes older fishing often becomes poorer not because there has necessarily been a decrease in primary production but because the community has be come diversified and tends to consume its own pro

... an eas-

...ishing

...g and

starting over or by other measures which accom plish the same thing ecologically With large res ervoirs the problem of maintaining continuous high yields of specific fish in the face of natural succes sion has not been solved *See* SUCCESSION ECOLOG ICAL

Species, individuals, and energy flow An im portant and little known area of ecology deals with the relationship between the number of species the number of individuals and the energy flow Most relatively stable natural communities contain a few species in each trophic group which are abundant and account for most of the energy flow Usually however the few common species are associated

with a large number of rare species The number of species relative to the number of individuals (species diversity) often increases with succession Because increasing diversity is not necessarily accompanied by increasing total productivity (in fact the reverse may be the case) it is now generally assumed that the advantage of diversity that is the survival value to the community lies in increased stability The more species present the greater are the possibilities for adaptation of the ecosystem to changing conditions whether these be short term or long term changes in climate or other factors Events which occurred in certain areas of Long Island Sound may be an example of such an adaptation which would not have been possible if rare species had not been present The development of large scale domestic duck farming on shore introduced large amounts of organic manure into the shallow waters The dominant or abundant phytoplankton producers of these waters were unable to tolerate the changed conditions but several other species which had formerly been very rare were able to tolerate and exploit the organic materials and soon became very abundant The productivity of the ecosystem was thus maintained (or actually increased because of the fertilization) because producers were present which could take over Unfortunately in this case the oysters could not use the new phytoplankton as food and the oyster industry (and its yield to man) was destroyed by the duck industry

As the human population of the world increases and more and more changes in the face of the earth are contemplated it is evident that increasing attention must be given to the total and not just the immediate effect of the changes It is probably desirable to maintain at least a moderate amount of diversity in nature because rare or so called useless organisms might some day prove to be very valuable In other words attention must be directed toward conservation of the ecosystem and not just conservation of individual organisms which are in demand at the moment

Primary production in the world The world distribution of primary production is shown schematically in Fig 3 Values represent the average gross production rate in grams of dry organic matter per square meter per day to be expected over an annual cycle For an estimate of total annual production multiply by 365 To visualize these values in terms of approximate kilocalories of potential food multiply by 5 As previously indicated as much as 90% of gross production may be available to heterotrophs but it should be remembered that man or any other single species cannot assimilate all energy fixed by plants For example corn stalks and wheat stubble and roots would be included in the total production of these crops but only the grain is currently consumed by man As may be seen from Fig 3 there are about three orders of magnitude in potential biological fertility of the world (1) some parts of the open oceans and land deserts which range around 0 1 g/(m²) (day) or less (2) semiarid grasslands coastal seas shallow lakes and ordinary agriculture which range between 1 and 10 (3) certain shallow water systems such as estuaries coral reefs and mineral springs together with moist forests intensive agriculture (such as year round culture of sugar cane or cropping on irrigated deserts) and natural communities on alluvial plains which may range from 10 to 20 g/(m²)(day) Production rates higher than 20 have been reported for experimental crops polluted waters and limited natural communities but these values are based on short term measurements values higher than 25 have not been obtained for extensive areas over long periods of time

Productivity controls Two tentative generalizations may be made from the data at hand First, basic primary productivity is not necessarily a function of the kind of producer organism or of the kind of medium (whether air fresh water or salt

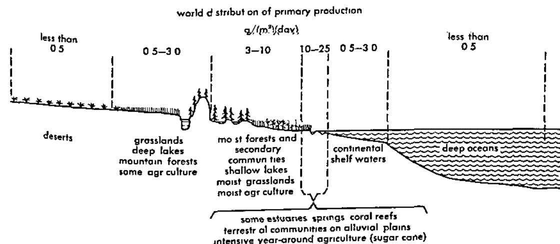


Fig 3 World distribution of primary production showing the range in gross production (as grams of dry matter per square meter per day) to be expected in the major environments of the world (*After E P Odum Fundamentals of Ecology 2d ed Saunders 1959*)

water) but is controlled by local supply of raw materials and solar energy and the ability of local communities as a whole (including man) to utilize and regenerate materials for continuous reuse. Terrestrial systems are not inherently different from aquatic situations if light, water and nutrient conditions are similar. However, large bodies of water are at a disadvantage because a large portion of light energy may be absorbed by the water before it reaches the site of photosynthesis. Second, a very large portion of the earth's surface is open ocean or arid and semiarid land and thus is in the low production category, because of lack of nutrients in the former and lack of water in the latter. Many deserts can be irrigated successfully and it is theoretically possible, and perhaps feasible in the future, to bring up nutrients from the bottom of the sea and thus greatly increase production. Such an upwelling occurs naturally in some coastal areas which have a productivity many times that of the average ocean.

Efficiency limits. It now seems clear that there is a rather definite upper limit to the efficiency with which light may be converted into organic matter on any large scale; this maximum has apparently been achieved by some natural communities (coral reefs, for example) as well as by the most efficient agriculture. In the former, of course, production is consumed by a large variety of organisms whereas ... f the net s temporar

grain production, for example, is 2 g/(m²)(day). The best immediate possibilities for increasing food production for man lie in measures which reduce ... he season of ... arge a ... RESH

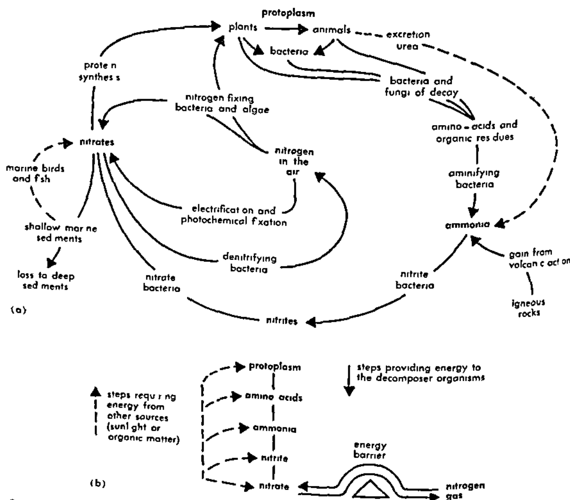
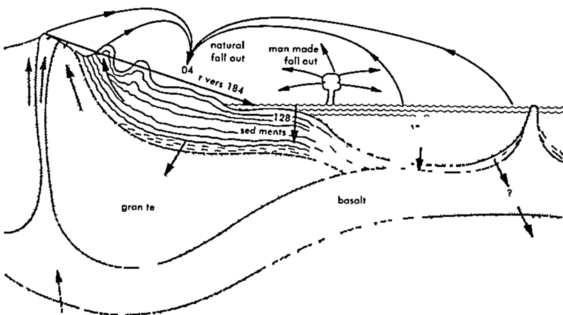


Fig. 4. Two ways of picturing the nitrogen biogeochemical cycle, a relatively perfect, self-regulating cycle in which there is little over-all change in available nitrogen in large ecosystems or in the biosphere as a whole, despite rapid circulation of materials. (a) Circulation of nitrogen between organisms and environment, depicted along with microorganisms which are responsible for key steps. (b) The same basic steps arranged in an ascending-descending series with the high-energy forms on top to distinguish steps which require energy from those which release energy. (After E. P. Odum, Fundamentals of Ecology, 2d ed., Saunders, 1959)



Fg 5 D agram of the sed mentary cycle involv ng movement of the more earthbound elements Where est mates are poss ble the amounts of material are es t mated n geograms/10^6 years (1 geogram = 10 ⁰ g)

WATER ECOSYSTEM, MARINE ECOSYSTEM TERRES TRIAL ECOSYSTEM

BIOGEOCHEMICAL CYCLES

The more or less circular paths of chemical ele ments passing back and forth between organisms and environment are known as biogeochemical cy cles The rate at which vital elements become avail able to biological components of the ecosystem is more important in determining primary and second ary productivity than flow of solar energy If an essential element or compound is in short supply in terms of potential growth the substance may be sa d to be a limiting factor The productivity of an entire ecosystem is sometimes limited by one ma terial available in least amount in terms of need accord ng to the principle of Liebig s law of the m n mum Thus water limits the desert ecosystem and n trogen or phosphorus often limits ocean eco sy tems However nature has considerable powers of adaptation and compensation In many environ ments species and varieties have evolved which have low requirements for scarce materials Also the amount of one substance which in itself may not be limiting often greatly affects the requirement for another su stance which is approaching a criti cal m nim m Conseq ently it is usually necessary to cons der the interaction of the essential materials if the limiting factors operating in a given situa t on are to be determined *See* BIOGEOCHEMISTRY

Nutrients Dissolved salts essential to life may be conveniently termed biogenic salts or nutrients They may be divided into two groups the macro nutrients and the micronutrients The macronutri ents include elements and their compounds needed in relatively large quantities for example carbon hydrogen oxygen nitrogen potassium cal cium phosphorus and magnesium The micronu trients include those elements and their compounds necessary for the operation of living systems but which are required only in minute amounts At least 10 micronutrients are known to be necessary for primary prod iction iron manganese copper zinc boron sodium molybdenum chlorine vanadium and cobalt Several others such as iodine are essen tial for certain heterotrophs Because minute re quirements are often associated with an equal or even greater minuteness in environmental occur rence the micronutrients deserve equal considera tion along with the macronutrients as possible lim iting factors

Types of cycle From the standpoint of the biosphere as a whole biogeochemical cycles fall into two groups the gaseous type cycles as ill is trated by the nitrogen cycle in Fig 4 and the sedi mentary type cycles illustrated in Fig 5 Cycles of oxygen carbon and water resemble the cycle of nitrogen in that a large gaseous pool is important in the continuous flow between inorganic and or ganic states As shown in Fig 4 the nitrogen of protoplasm is broken down from organic to inor ganic form by a series of decomposer bacteria each specialized for a particular part of the job The ni trogen ends up as nitrate or other form usable by

green plants in the synthesis of new organic matter The air is the great reservoir and safety valve of the system Nitrogen is continually entering the air by action of denitrifying bacteria and continually returning to the cycle through the action of nitrogen fixing organisms and also through nonbiological fixation Only certain bacteria and algae (which however are abundant in water and on land) can fix nitrogen No so called higher plant or animal has this ability legumes fix nitrogen only because of the symbiotic bacteria which live in their roots The steps from protoplasm down to nitrates provide energy for the decomposer organisms whereas the return steps require energy from other sources such as from organic matter or sunlight Likewise nitrogen fixers must use up some of their carbohydrate or other energy stores in order to transform atmospheric nitrogen into nitrates

Feedback mechanisms The self regulating feedback mechanisms as shown in a simplified manner in Fig 4 make the nitrogen and the other gaseous type cycles relatively perfect when large areas of the biosphere are considered Thus increased movement of materials along one path is quickly compensated for by adjustments along other paths However nitrogen often becomes a limiting factor locally either because regeneration is too slow or because loss from the local system is too rapid

Most biogenic substances are more earthbound than nitrogen and their cycles follow the pattern of erosion sedimentation mountain building and volcanic activity as shown in Fig 5 Biological activity on land and in the upper layers of water results in local cycles from which there is usually a continual loss downhill and replacement from up h ll

good example Phosphorus is relatively rare in the surface materials of the

all is being replaced by natural processes For the time being man is able to mine the considerable reserves of underground phosphate rock and make up some of this loss but eventually a means may have to be found to recover phosphorus from the sea

Nonessential and radioactive elements Biogeochemical cycles involve elements essential to life The nonessential elements pass back and forth between organisms and environment and many of them

as

val

come concentrated in tissues apparently because of similarity to specific vital elements The ecologist would have little interest in most of the nonessential elements were it not for the fact that atomic bombs and nuclear power operations produce radioactive isotopes of some of these elements which then find their way into the environment and into food chains Even a rare element in the form of a radioactive isotope can be of biological concern because a very small amount of material from a geochemical standpoint can have marked biological effects Thus the cycling of such things as strontium cesium cerium ruthenium and many others may be of great concern in coming years At present strontium is receiving special attention Strontium behaves like calcium and follows it into biological systems None of the strontium which is naturally involved in the calcium cycle is radioactive but radioactive strontium is now becoming widespread in the biosphere as a result of fallout from atomic weapons tests Small amounts of radiostrontium have now followed calcium from soil and water into vegetation animals milk and other human food and human bones in almost all parts of the world In 1958 several hundred micromicrocuries (1 micromicrocurie = 10^{-12} curies) of radiostrontium were present for every gram of calcium in some soils The bones of children in North America and Europe averaged 1 2 micromicrocuries g of calcium There is at present considerable controversy as to whether these small amounts are detrimental but most scientists agree that it would be desirable to keep as much radioactivity out of the food chain as possible

Tracers There is a bright side to the atomic age production of radioactive isotopes Tiny amounts of such isotopes provide convenient tracers whereby the movement of materials in ecosystems can be accurately measured Much has already been learned about the cycling and turnover rates of phosphorus through the use of the isotope P^{32} The radioactive isotope of carbon (C^{14}) has proved invaluable in measuring the rate of primary production in the ocean It is certain that intelligent use of radioisotopes as tools can help solve problems of the atomic age [E P O]

Bibliography W C Allee et al *Principles of Animal Ecology* 1949 E P Odum *Fundamentals of Ecology* 2d ed 1959 H J Oosting *The Study of Plant Communities* 2d ed 1956

Ecology, applied

Mankind is changing the face of the earth to such an extent that many of the present plant and animal communities and even the land surfaces themselves bear the marks of his interference Applied ecology deals with man s activities in managing natural resources a management based upon a knowledge of basic ecology and upon how man s efforts can change actions reactions and interactions with the communities of plants and animals An ecological approach is implicit in all agriculture forestry range management wildlife management fisheries management and other aspects of natural resource man logy is usually

cifi

ncy

e

successful applications of ec knowledge of the life histories of the plants and animals to be managed the effect of environments

upon these life histories the interactions of the plants and animals making up the community and the means at the disposal of man to change these relationships These applications may be illustrated for the major fields of natural resource management

Forestry. The management of forests for man's own use through silviculture provides a prime example of applied ecology Forest communities are usually much closer to natural communities than agricultural crop communities so that the relationship is clear between natural ecological processes and these processes as affected by man's activities

Classically forestry is oriented toward the growing of sawtimber as a source of lumber plywood and veneers In the twentieth century however the rapid development of the paper and plastics industries has resulted in increasing emphasis upon growing trees as a source of fibers and raw materials for chemical products especially cellulose During the same period the concept of multiple use of forests has gained increasing acceptance so that management for compatible uses including wood production, wildlife production grazing recreation and watershed protection is being increasingly emphasized

Occasionally it may be to man's interest to perpetuate climax communities which develop over a long period of time in the absence of severe disturbances (*see* CLIMAX COMMUNITY) The hemlock beech sugar maple forest in the northeastern United States and the Norway spruce-silver fir beech forest in central Europe are examples of economically valuable climax forest associations More often than not however the valuable forest types will be replaced in time by less valuable forest types unless man intervenes Thus the white and red pine forests of New England and the Lake States will be replaced in the absence of disturbance by shade tolerant hemlock and hardwoods which develop as an understory under the aging pine Similarly the pines of the southeastern United States are subclimax to mixed hardwoods on most sites and Douglas fir in the Pacific Northwest is usually subclimax to western hemlock western red cedar and other shade tolerant species *See* ANIMAL COMMUNITY PLANT COMMUNITY

To encourage the desired tree species and to obtain the desired stand structure the silviculturist employs cutting planting fire chemicals and mechanical treatment By logging the existing stand in a patterned partial cutting or by clearcutting environmental conditions may be created that will favor the desired species If natural seeding is not successful the desired species may be planted Unwanted species may be killed or discouraged by prescribed burning or by chemicals Mechanical scarification of the site combines brush control with the preparation of the soil for seedling establishment

Once the forest is established it must be kept vigorous and healthy A knowledge of growing space requirements for individual trees is used as a basis for systematic weedings and thinnings Forest insects and diseases must be held in check largely through cultural and biological control by the establishment of conditions unfavorable to pests or favorable to their parasites and predators Direct control of forest pests is usually uneconomic because forests are relatively low in value per unit of area *See* ENTOMOLOGY ECONOMIC

The success of a silvicultural treatment depends largely upon how well the silviculturist understands the ecology of the particular forest community being managed and upon the timing and intensity of his treatment *See* CLIMAX PLANT FORMATIONS FOREST ECOLOGY



(... Olympic Nations ... orest Service)

Agriculture The artificial or man made nature of most agricultural crop and domestic animal communities tends to obscure the ecological nature of agriculture The requirements of a given economic plant (such as corn cotton or coconut) or animal (cattle sheep chickens) as to climate soil and other growing conditions have been determined over the ages by trial and error and are so well known as to be axiomatic The present day production of many new hybrids and other products of plant and animal breeding however constantly poses the problem of what the environmental requirements of these new organisms are and where and how they can be introduced into existing plant and animal communities *See* AGRICULTURE

The effect of environment upon plant and animal growth is being studied increasingly in controlled environmental chambers where plants and animals can be subjected to specified day lengths temperatures humidities and other aspects of the environment The phytotron a series of air conditioned greenhouses for studying plants at the California Institute of Technology is a well known and unusually elaborate example of such chambers

Of critical importance to continued agriculture is the prevention of soil erosion and of chemical depletion of the soil upon which crops are grown or on which farm animals are reared The soil constitutes an ecological system formed of a complex of minerals organic residues and living plants and animals Only through the application of ecological principles can this complex be maintained in such a condition as to permit the indefinite use of land for the maximum production of agricultural products The effects [illegible] crops [illegible] tions [illegible] can [illegible] the soil ecosystem as a whole is clearly comprehended

Range management As an aspect of applied ecology range management represents more intensive effort on the part of man than most forestry activities but less than most cropland practices The open grasslands and savanna woodlands of the American West and many other parts of the world can support large numbers of cattle sheep and other grazing animals only if the food and water supply of the range is maintained Grasslands may readily be overgrazed with resultant development of less palatable communities soil erosion and even destruction of the site The maintenance of the proper number and kind of stock at the various times of the year is the basis of modern range management Uniform use of the available range can be approached through the development of water supplies the placing of salt licks and the forced movement of the stock *See* RANGE LAND CONSERVATION

Wildlife management The scientific production of deer grouse quail ducks and other desired wildlife is a rapidly growing aspect of applied ecology The near extermination of wolf puma and other natural predators the heavy pressures of hunting by man which have greatly reduced populations of bison elk caribou, and other game animals and the change in ecological balances produced by the introduction of exotic animals and plants such as the pheasant in the northern United States and the eucalyptus in California have all changed the numbers and distribution of game animals, both mammals and birds Maximum healthy populations of desired animals can be maintained by creating optimum environments mainly through the development of increased food supplies and increased shelter Also through controlled hunting and through management of natural predators, populations can be regulated at a level that will permit the appropriate number of animals to weather unfavorable seasons and at the same time minimize losses through starvation and disease *See* WILDLIFE CONSERVATION

Fisheries management The management of fisheries for food or sport provides another example of applied ecology In contained bodies such as ponds lakes reservoirs and rivers a fair degree of population and environmental control may be possible Bodies of water may be drained undesirable fish may be trapped poisoned or otherwise removed desirable fish may be introduced food plants and animals may be added and even fertilizers may be spread Here again the maintenance of the wanted fish at the highest levels is the desired end and this can be achieved only through understanding the community as a whole and managing it as a whole In the oceans and other large bodies of water management of the whole community is seldom practicable but certain aspects of the life cycle of certain commercial fishes can be regulated through controlled fishing and management of spawning areas The development of favorable environmental conditions for the spawning of salmon in rivers and streams and the creation of unfavorable conditions for the spawning of the lamprey in the Great Lakes are examples of applied ecology in commercial fisheries management *See* FISHERIES CONSERVATION

Water resource management The production of clear unpolluted water for human use is fast becoming a critical problem in many areas of the United States and has long been a controlling factor in determining the distribution of human populations In the hydrologic cycle much of the water that falls as precipitation is returned to the atmosphere through evaporation and through transpiration by plants Through changing the nature and distribution of the plant cover, changes which lead to changes in the soil structure the amount of water that enters the ground can be increased The hydrologic management of streams through such devices as channel improvement dams and canals can bring larger quantities of clearer water to points of human use whether for agriculture industry or human consumption Finally control of stream pollution through regulation of the source of impurities filtering and chemical treatment represents applications of ecology that increase

available water supply both to man and to other animals See ECOLOGY HUMAN WATER CONSERVATION [S H S]

Bibliography E P Odum *Fundamentals of Ecology* 1953 H J Oosting *The Study of Plant Communities* 2d ed 1956 *Symposium on applied ecology Ecology*, 38 46–64 1957

Ecology, human

That subdivision of ecology which is concerned with the relations between man and his environment Human environments are composed of the physical features of the habitat the plants and animals associated with man and other human beings Physical factors important to man include the topography soil climate useful minerals and radiations Certain plant and animal associates of man provide him directly or indirectly with food clothing and shelter and may furnish materials for industries of various kinds Injurious forms may attack man himself or may compete with or damage cultivated crops domestic animals or valuable wild species The interrelations among the numerous plant and animal species that compose any ecological community of which man is a member are always complex The extreme of complexity is reached in a modern industrial state

In order to continue to exist each human being must have a constant supply of air for breathing be protected against extremes of heat or cold escape serious infection by disease organisms and parasites and avoid being injured or eaten by large predators Regulatory mechanisms within the body maintain a suitable balance (homeostasis) of oxygen content in the tissues carbon dioxide and other excretory products water content salt content nutritive elements temperature and all the other factors that are essential to the proper functioning of the body The sense organs are used to detect food enemies friends and conditions suitable for survival and bodily welfare The muscular and locomotor systems enable the individual to seek out favorable situations and to secure the needed food clothing and shelter Each individual thus constantly maintains a dynamic equilibrium between its internal physiology and his external environment Parts of these ecological relations of the human individual to his personal environment are usually treated under the science of physiology and other parts under psychology See HOMEOSTASIS

By the evolution of intelligence and reason and by the acquisition of culture man has greatly increased his ability to modify his immediate environment and has tremendously expanded his physical and biotic resources Social cooperation is a key feature of this evolution A small group of persons may live in relative comfort under circumstances where a single individual would have difficulty in surviving

Cultural evolution Human social groups evolve cultures which are adapted at least tolerably to the climate and other physical conditions the minerals the native and domesticated plants and animals and the diseases of their local habitats Human ecology therefore involves features of human relations which are treated by anthropology botany climatology demography economics geography geology sociology and zoology Human history has been to a considerable degree affected by the resources available to tribes nations and empires and also by their climates and routes of communication Many branches of social science thus have an ecological basis All the branches of applied biology may likewise be considered to have an ecological foundation Here may be included agriculture animal husbandry conservation forestry game and fish management and human hygiene

Behavior Human behavior is another important aspect of human ecology The customs and ideals of individuals and groups are affected by their physical biotic and social environments Each individual and each social group must have methods of response to its environment which are sufficiently effective to insure survival Every social group likewise has customs which govern the relations among the individuals within the group and with strangers The elements of appropriate behavior are inculcated in each individual from early infancy onward and are maintained by pressure of public opinion Variability in behavior occurs in every population and provides the possibility of further evolution in behavior type Studies of human behavior have mostly been made by persons trained in anthropology psychology or sociology

Human society Any study in human ecology is usually focused on a particular region and its human society Such a survey may describe the local physical conditions including topography soils and climate certain species of the fauna and flora with special attention to those forms most important to man the native vegetation cultivated crops domestic animals useful minerals and articles obtained by trade The people and their culture whether primitive or modern may be described The numbers local distribution and the age and sex composition of the population may be estimated The organization of the human society according to race language religion occupation and social and economic class may be considered The history of the human occupancy is also often discussed The most important part of any such

of interdependence of the several economic social and political classes regulatory mechanisms that keep the community adjusted to the vicissitudes of the habitat and to internal changes in the community itself relations of the community to other adjacent or distant human communities degree of utilization and conservation of the natural resources and trends in population density population quality economic resources and

social organization Most regional surveys of course consider only a few of the topics listed above

Human relations Although a large share of human activities thus has an ecological basis no one person can possibly be competent in this entire field of knowledge Critical investigation of the relations of a human individual or group to any type of environment nearly always requires the cooperation of persons trained in several different disciplines Human ecology consequently cannot properly be assigned to any one of the well established departments of knowledge A separate science of human ecology however has not as yet evolved Practical quantitative methods for measuring or evaluating the relations between human individuals or between societies and their physical or biotic environments are largely lacking The concepts [illegible]

knowledge which includes all those branches of scientific inquiry which are concerned in any way with the relations between human activities and environmental factors *See* COMMUNITY ECOLOGY ECOLOGY APPLIED ENVIRONMENT *see also* PSYCHOLOGY PHYSIOLOGICAL AND EXPERIMENTAL

[illegible] [L R D]

[illegible] M Hauser and O D Duncan (eds) *The Study of Population* 1959 A W Hawley *Human Ecology* 1950 J A Quinn *Human Ecology* 1950 C L White and G T Renner *Human Geography* 1918 G K Zipf *Human Behavior and the Principle of Least Effort* 1949

Ecology, physiological

The study of plant processes under natural or simulated environmental conditions (*see* ENVIRONMENT) Some of these processes are germination growth photosynthesis respiration absorption translocation transpiration and reproduction Each of these is a complex series of physical and chemical reactions in cells and tissues of the plant

Scope Physiological ecology attempts to explain why a certain species of plant can grow in a particular environment This question has not been answered completely for any species Any answer requires a knowledge of the genetic structure of the species and the rates of the physiological processes in its individuals as affected by the whole complex of environmental components

The potential geographic range of a species both natural and cultivated depends upon the nature and degree of genetic variation among its individuals and the frequency and extent of occurrence of suitable environments These two complex variables operating together determine the rates of physiological processes and thus the degree of success in growth and reproduction of individuals of a species in a particular environment

Widespread species are genetically diverse Within such species ecological races have evolved in response to environmental selection of genetic material These races consist of local populations with enough genetic diversity to allow for some survival within certain limits of environmental change Physiological ecological studies of process rates within local populations and ecological races provide a measure of the environmental tolerance limits of a species Such a comparative physiological ecology is a relatively new interdisciplinary field and the data which it can provide are not yet abundant

Such data may be obtained in natural environments under field conditions or in controlled simulated environments Field measurements are under the full impact of the uncontrolled environment, but have the advantage of providing realistic values Measurements under controlled laboratory conditions are easier to make more precise but the technical difficulties of exact reproduction of many kinds of natural environments are great and costly

Field measurements. Genetically determined ecological races are found by making transplants of individuals of a species from contrasting environments to uniform gardens In these transplant gardens the growth and reproduction of members of each local population may be measured and compared in correlation with uncontrolled environmental cycles In this way, many species have been found to consist of climatic edaphic and biotic races or ecotypes

The physiological processes in individuals of ecotypes can be measured in their natural habitats if portable equipment is available The processes most often studied are transpiration photosynthesis and respiration

Transpiration rates of nonwoody plants or young trees or shrubs may be measured by loss of weight of plants with their roots in sealed containers (phytometers) or with the plant enclosed in an air stream chamber using a portable infrared gas analyzer adapted for the absorption spectrum of water vapor or as an alternative by the use of electrical humidity sensing devices Transpiration rates of larger woody plants are difficult to measure directly Immersing the cut end of a twig in water in a horizontal buretlike potometer measures transpiration indirectly through absorption Another method involves weighing a freshly cut twig and reweighing after a few minutes transpiration is equal to loss in weight per unit of time Transpiration rates must be related to external leaf surface area internal leaf surface or number of stomates

Apparent photosynthesis may be measured by enclosing the plant or twig in a sealed transparent chamber passing air through the chamber and measuring decrease in carbon dioxide (CO_2) by absorbing CO_2 in potassium hydroxide solution and titrating A better method employs an infrared gas analyzer adapted for the CO_2 absorption spectrum The decreased CO_2 content of the air after passing through the chamber can be continuously

Portable infrared gas analyzer being used to measure photosynthesis of alpine plants at 11 000 ft in Wyoming The plant is in the small chamber in the right background and is connected by tubing to the pumping and flow-meter apparatus at the left, then to the analyzer amplifier, and recorder on the right

recorded in either a closed or open system (see illustration) Overheating of the plant chamber by solar radiation ... der re... tosynt... to chl... contents or both Because photosynthesis is the basic process in ecosystem dynamics such field photosynthesis rate measurements are fundamental to measurements of community productivity

Respiration may be measured by either of the two above methods by darkening the plant chamber and measuring CO_2 production

Controlled environment measurements All processes may be measured with greater ease precision and environmental variety in controlled environment facilities The principal disadvantages of such facilities are difficulties of exactly reproducing natural environments even in a small space and their cost

Two principal types of controlled environment facilities are in use (1) relatively small reach in or walk in chambers in which thermoperiod photoperiod light intensity, and sometimes humidity are controlled and (2) phytotrons relatively large air conditioned buildings in which a variety of controlled environments are available Until 1958 there were only two phytotron... nal at the E... the Californi... at Wageningen Netherlands Others are under construction including a large one near Gif sur Yvette France

Small controlled chambers and phytotrons are in valuable for work in physiological ecology because the environments in them can be programmed and changed at will The principal problems in physical environmental control are obtaining a light intensity approaching that of natural sunlight over considerable area without overheating and good control of humidity

One important contribution of phytotron research is the evidence for much better growth of most plants at strongly different photo and nycto-temperatures rather than at the classical constant temperatures of most laboratory experiments This is in agreement with growth and temperature cycles under natural conditions

Dependent upon the development of physiological ecology are knowledge of tolerance limits of ... portable and precise field equipment and methods and greater available space in versatile controlled condition laboratories See PLANT METABOLISM PLANT PHYSIOLOGY [W D B]

Bibliography W D B... ecology *Ann* ... J C Clausen ... *cies* 1951 J ... and W M Hiesey *Experimental Studies on the Nature of Species IV, Genetic Structure of Ecological Races*, Carnegie Inst Wash Publ 615 1958 H R Highkin Temperature-induced variability in peas *Am J Botany*, 45 626–631 1958 P J Kramer Some effects of various combinations of day and night temperatures and photoperiod on the height growth of loblolly pine seedlings *Forest Sci* 3(1) 45–55 1957 F W Went *The Experimental Control of Plant Growth* 1957

Economizer and air heater

Components of a steam generating unit that are capable of absorbing heat from the products of combustion after they have been cooled below the temperature required for the generation of steam

The economizer is supplied by feed water at a temperature below saturation and delivers water at higher temperature to the boiler drum It is constructed of pressure tubes arranged as a convection bank Size of an economizer in proportion to that

gas outlet
air inlet
air by pass
air
air
air outlet
gas inlet
gas upflow
air counterflow three pass

Fig 1 Recuperative air heater

social organization. Most regional surveys of course consider only a few of the topics listed above.

Human relations. Although a large share of human activities thus has an ecological basis, no one person can possibly be competent in this entire field of knowledge. Critical investigation of the relations of a human individual or group to any type of environment nearly always requires the cooperation of persons trained in several different disciplines. Human ecology consequently cannot properly be assigned to any one of the well-established departments of knowledge. A separate science of human ecology, however, has not as yet evolved. Practical quantitative methods for measuring or evaluating the relations between human individuals or between societies and their physical or biotic environments are largely lacking. The concepts of human ecology are mostly those of the individual disciplines involved. Human ecology, therefore, may be considered to be a broad field of knowledge which includes all those branches of scientific inquiry which are concerned in any way with the relations between human activities and environmental factors. See COMMUNITY; ECOLOGY; ECOLOGY, APPLIED; ENVIRONMENT; see also PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL.

[L.R.D.]

Bibliography: J. W. Bews, *Human Ecology*, 1935; F. F. Darling (ed.), *West Highland Survey*, 1955; L. R. Dice, *Man's Nature and Nature's Man*, 1955; P. M. Hauser and O. D. Duncan (eds.), *The Study of Population*, 1959; A. W. Hawley, *Human Ecology*, 1950; J. A. Quinn, *Human Ecology*, 1950; C. L. White and G. T. Renner, *Human Geography*, 1948; G. K. Zipf, *Human Behavior and the Principle of Least Effort*, 1949.

Ecology, physiological

The study of plant processes under natural or simulated environmental conditions (see ENVIRONMENT). Some of these processes are germination, growth, photosynthesis, respiration, absorption, translocation, transpiration, and reproduction. Each of these is a complex series of physical and chemical reactions in cells and tissues of the plant.

Scope. Physiological ecology attempts to explain why a certain species of plant can grow in a particular environment. This question has not been answered completely for any species. Any answer requires a knowledge of the genetic structure of the species and the rates of the physiological processes in its individuals, as affected by the whole complex of environmental components.

The potential geographic range of a species, both natural and cultivated, depends upon the nature and degree of genetic variation among its individuals and the frequency and extent of occurrence of suitable environments. These two complex variables, operating together, determine the rates of physiological processes and thus the degree of success in growth and reproduction of individuals of a species in a particular environment.

Widespread species are genetically diverse. Within such species, ecological races have evolved in response to environmental selection of genetic material. These races consist of local populations with enough genetic diversity to allow for some survival within certain limits of environmental change. Physiological-ecological studies of process rates within local populations and ecological races provide a measure of the environmental tolerance limits of a species. Such a comparative physiological ecology is a relatively new interdisciplinary field, and the data which it can provide are not yet abundant.

Such data may be obtained in natural environments under field conditions or in controlled simulated environments. Field measurements are under the full impact of the uncontrolled environment but have the advantage of providing realistic values. Measurements under controlled laboratory conditions are easier to make more precise, but the technical difficulties of exact reproduction of many kinds of natural environments are great and costly.

Field measurements. Genetically determined ecological races are found by making transplants of individuals of a species from contrasting environments to uniform gardens. In these transplant gardens the growth and reproduction of members of each local population may be measured and compared in correlation with uncontrolled environmental cycles. In this way, many species have been found to consist of climatic, edaphic, and biotic races or ecotypes.

The physiological processes in individuals of ecotypes can be measured in their natural habitats if portable equipment is available. The processes most often studied are transpiration, photosynthesis, and respiration.

Transpiration rates of nonwoody plants or young trees or shrubs may be measured by loss of weight of plants with their roots in sealed containers (phytometers), or, with the plant enclosed in an air-stream chamber, using a portable infrared gas analyzer adapted for the absorption spectrum of water vapor or, as an alternative, by the use of electrical humidity-sensing devices. Transpiration rates of larger woody plants are difficult to measure directly. Immersing the cut end of a twig in water in a horizontal buretlike potometer measures transpiration indirectly through absorption. Another method involves weighing a freshly cut twig and reweighing after a few minutes, transpiration being equal to loss in weight per unit of time. Transpiration rates must be related to external leaf surface area, internal leaf surface, or number of stomates.

Apparent photosynthesis may be measured by enclosing the plant or twig in a sealed transparent chamber, passing air through the chamber, and measuring decrease in carbon dioxide (CO_2) by absorbing CO_2 in potassium hydroxide solution and titrating. A better method employs an infrared gas analyzer adapted for the CO_2 absorption spectrum. The decreased CO_2 content of the air after passing through the chamber can be continuously

ties and environments of the open water those of the shores and those of the deeper bottom (*see* FRESH WATER ECOSYSTEM MARINE ECOSYSTEM)

Some major features of ecosystems important in ecological understanding are discussed here

Complexity and interrelations The scientist studying an ecosystem is confronted by a bewildering variety of interrelations The environmental factors which can be specified and measured are variously interrelated in their effects on one another and in their significance to the physiology of the organisms Thus the effects of humidity cannot simply be measured by relative humidity for its effect on any given organism is conditioned also by the existing temperature and wind velocity by the position in the community and characteristics of the organism concerned and by the results of water loss on the physiology of that organism The environment affecting the organisms in a community is often conceived as the environmental complex or the sum of all distinguishable environmental factors together with the interrelations among these The factors may in fact be regarded as isolated from the whole environmental complex consisting of those particular features abstracted from the whole and chosen for measurement and study Different organisms in the community are variously affected by different factors and combinations of factors Different organisms are also most variously affected by major kinds of interrelations such as food relations shelter competition for space and other resources of environment shading chemical influences and others The sum of these interrelations among species in the community which so link together the species that almost any one may directly or indirectly affect almost any other is referred to as the web of life The niche of a species may be considered with respect to its position in the web of life that is in relation to other species and the community as a whole Such is the complexity of an ecosystem that it is generally not possible to study all its features at once or to seek *complete understanding of all its interrelations* (*see* ECOLOGIC INTERACTIONS)

Environmental dependence and modification Environment and community are always intimately interrelated The community is necessarily dependent on environment and cannot be understood apart from it The relation between environment and community is not one of simple cause and effect however for environmental factors are in various ways and to varying degrees modified by the community Some factors such as salinity of sea water temperature of water of a pond and others may be scarcely affected by the presence of organisms others such as humidity and wind velocity inside a forest and concentration of phosphates in the water are strongly modified by organisms or are determined by the activities of organisms Still others for example organic matter in water and soil exist because of the activities of organisms Some communities such as a highly developed stable forest modify the environments in which their component organisms live more than other communities such as those of deserts and the first stages of successions In general however there is a pervasive reciprocity between environment and community in determining the characteristics of both environment and community

Community expression of environment Characteristics of environment are reflected in characteristics of the community The community may be said to express its environment Certain broad correlations between kinds of environments and kinds of communities around the world can be observed Among terrestrial communities climate is expressed in the over all structure and composition in terms of major kinds of plants the physiognomy of vegetation (*see* TERRESTRIAL ECOSYSTEM) Thus, wherever tropical climates with high rainfall throughout the year occur these climates support the kind of community known as a tropical rain forest Wherever temperate continental climates with sufficient summer rainfall occur these support the temperate deciduous forest One may thus recognize the general biological phenomenon of adaptation to environment on two levels first the organism which must have structure and function suited to survival in its niche and the environmental complex of its ecosystem and second the community which must have structure and function appropriate to utilization of the resources of and persistence in its environment *See* CLIMAX PLANT FORMATIONS

Spatial patterning Many environmental factors form gradients in space as they are followed from one point to another on the earth s surface The many factors of the environment change in different directions through space some in correlation with one another others in partial or complete independence Environments consequently form in space a complex pattern of factor gradients At each point in this environmental pattern a natural community develops in dependence upon the environment of that point To the pattern of environments in space there consequently corresponds a pattern of communities and the populations and other characteristics of communities form complex patterns of gradients in space as do the factors of environment Often this patterning in space has a self repeating character as in an area of low mountains in New York where one may find a pattern of hemlock forests in ravines and north facing slopes oak forests on most other slopes and grassy oak openings on dry south slopes with this pattern repeating itself across a series of valleys and ridges Environments and communities in general grade continuously into one another along spatial gradients of environment Ecosystems and communities consequently do not like organisms form distinct clearly bounded individuals separate from each other There are exceptions to this A lake may be regarded as an ecosystem with a fairly clear bound

ary. Relative discontinuities between communities occur because of disturbance and where environment is relatively discontinuous, as at the shores of lakes and seas. When such relative discontinuities or ecotones are studied, they are often found to represent not merely the meeting place of two communities but distinctive communities in their own rights.

Rhythms. Environments vary also in time, in part irregularly but more generally and more significantly in regular periodic rhythms. Only the abyssal communities of the ocean depths live in a constant environment. Other communities are subject to annual and diurnal cycles, and communities of the seashore to more complex rhythms of tides. Organisms of the community respond in various ways in their cycles of activity to these rhythms of environment. Consequently the environmental rhythms impose self-repeating patterns of change in time on the community. Similar activities may be pursued by different species at different times in these cycles. Thus flycatchers feeding in the daytime on diurnal insects may be replaced at dusk by nighthawks feeding on nocturnal insects; spring flowers and the insects visiting them may be replaced in fall by quite different groups of flowers and insect visitors. Daily and yearly rhythms thereby permit specialization by time of function among the species of a community and differentiation in time of the community as a whole.

Community differentiation. A natural community comprises a large number of species, but in general no two species are exact competitors. The species of the community are specialized to fill different though often overlapping niches. Species with similar niche requirements may be active at different times or in different places in the community. The specialization of species and differentiation of the community are illustrated in the stratification of terrestrial communities. A forest may include five or more strata, the upper and lower tree layers, shrub, herb, and moss layers, with each stratum having its characteristic plant and animal species. Parallel activities may be pursued by different species in different strata. For example, one bird species may nest and feed on insects in the tree stratum, another on the ground. Such vertical differentiation in space is a basis of one of the major characteristics of communities—their species diversity, or relative richness in numbers of species. One may generalize that species diversity, differentiation into strata, and productivity of terrestrial communities are maximal in most favorable environments, such as that of the tropical rainforest, and decrease toward environments which are unfavorable or extreme. There are, however, various limitations to this statement, and species diversity and productivity do not simply parallel one another.

Productivity relations. Productivity is even more than species diversity a fundamental characteristic of communities and ecosystems. It is perhaps the most important single feature of the community.

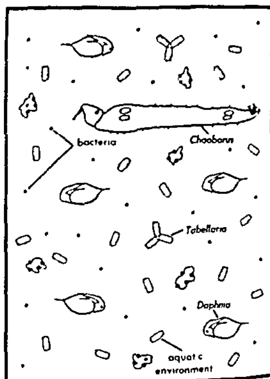


Fig. 1. A much simplified ecosystem consisting of a small volume of water and plankton organisms (microorganisms which float in the water). Components of the ecosystem in this illustration are the environment, the water and its dissolved material; green plants or producers, shown as the rectangular cells (*Tabellaria*, a diatom); herbivorous animals or primary consumers (*Daphnia*, a water flea); carnivorous animals or secondary consumers (*Chaoborus*, a fly larva); and bacteria or reducers floating in water or acting on irregular masses of decomposing organic matter. The various organisms are not drawn in correct size relations.

Productivity may be defined as the amount of energy or matter bound into organic compounds by organisms per year per unit of the earth's surface. Although particular environmental factors may effectively limit productivity, it is in general a complex resultant of all factors of the environment, an expression in biological activity of the environmental resources and of other factors affecting utilization of these resources by the community. Major factors affecting productivity include those of temperature, nutrient availability, and water availability. Certain broad correlations of productivity with environments may be observed, such as the decline along the temperature gradient from tropical rainforest to the treeless arctic tundra and along the moisture gradient from tropical rainforest to desert. Productivity thus underlies such other characteristics of communities as their structure or physiognomy and stratification (see BIOLOGICAL PRODUCTIVITY).

Food chains, trophic levels, and pyramids. Productivity of a community is based wholly on the activities of green plants, except for certain

autotrophic bacteria and the communities of such lightless environments as the depths of oceans and lakes, which are dependent on other communities for their food Green plants use the energy of sun light to produce organic substances from carbon dioxide and water These plants are then eaten by animals and these animals by other animals In the community illustrated in Fig 1 organic mat ter and food energy might be passed along from the diatoms to the water fleas to the *Chaoborus* larva and from this to a small fish and to a heron Such a sequence of organisms is referred to as a food chain Since many species feed on a number of other species food chains as they are variously interrelated form an important part of the web of life in a community Certain major steps or trophic levels in food chains may be recognized Among these are the green plants or producers the herbivorous animals or primary consumers the carnivorous animals or secondary consumers and carnivorous animals feeding on carnivorous animals or tertiary consumers At each level most of the energy available to organisms is expended in the life activities of those organisms Only a fraction of this energy can be harvested and used by the next trophic level There is consequently a stepwise decrease of productivity through the se quence of trophic levels This relation is known as the pyramid of life

Cycling and the functional kingdoms Food chains and pyramids are part of a broader phe nomenon the cycling of materials between organ isms and environment in the ecosystem A given substance such as phosphate in the soil may be taken up by the roots of green plants and used in organic syntheses and then passed along food chains through animals until with the death and decomposition of these organisms the phosphate is released and returned to the soil where it is again available to green plants Decomposition of organic material results largely from the activities of certain organisms such as bacteria and fungi One may thus recognize three major nutritional groupings and evolutionary directions as the func tional kingdoms' in communities These are the true or green plants or producers characterized by photosynthesis as their mode of nutrition the

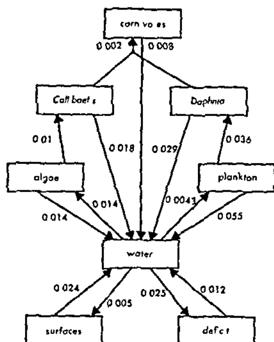


Fig 3 A simplified pattern of phosphate movement in a pond based on an experiment by R H Whittaker with radioactive phosphorus (P) tracer Numbers are transfer rates (that fraction of P^{32} in the box at the tail of the arrow moving in the direction indicated per hour) Amount of P^{32} in water at a given time depends on the various rate values and the temporary steady-state of P^{32} distribution between water and organisms Organisms indicated include green plants—plankton the floating algal cells and algae attached algae of rock surfaces primary consumers—Daphnia a water flea and Callibaetis a mayfly nymph and secondary consumers—carnivores water boatmen dragonfly nymphs small fish and so forth Surface is the film of bacteria and other microorganisms on rock surfaces; deficit to the P^{32} mainly in the depths of the rocky substrate not otherwise accounted for

animals or consumers characterized by consump tion or ingestion of organic food and the bacteria

these in cycling of materials through the ecosys tem is illustrated in Fig 2 When the movement of particular substances is studied in detail a great complexity of routes of movement in the ecosys tem may be revealed This is best known for the nitrogen cycle Concentrations of many substances in the environment are determined by the cycling of the material through the community (Fig 3) Productivity relations also are determined not simply by the amounts of nutrients and trial communities the amount of wa vironment but by the rate and r

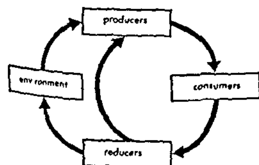


Fig 2 The generalized cycle of materials in an eco system

Energy flow and steady state conditions Energy also flows in a cycle of uptake from the environment movement through the community and dissipation back to the environment The community and the ecosystem are thus like organisms open energy systems In a stable ecosystem the community possesses a stable pool of available energy of organic compounds and the flow of energy into this pool by photosynthesis is balanced by the outward flow by dissipation which returns energy to the environment With regard to both energy and matter the community is in steady state a dynamic equilibrium of apparent constancy underlying which there is a continuing flow of energy and matter As an open energy system in steady state the community like the organism resists running down to maximum entropy according to the Second Law of Thermodynamics by its continuing intake of energy and passing of this energy through the trophic levels as the negative entropy of food Not only the community as a whole and its different trophic levels but the species populations of a stable community are in steady state The steady state for the species involves a balance of individual births and deaths while the population itself remains essentially stable or fluctuates around a stable average See POPULATION DYNAMICS

Succession Not all communities are fully stabilized in this sense and many become so only after an extended process of community development or succession If a bare area of the earth s surface is exposed as by a landslide this bare area may be occupied first by such simple plants as lichens and mosses These plants contribute along with physical processes to the breakdown of the rock and formation of soil until higher plants such as grasses may occupy the environment The grasses may in turn make it possible for shrubs to grow and the shrubs make growth possible for trees until finally a stable forest may result Through the course of succession there tends to be increasing productivity species diversity and stability of the community with increasing development of the soil increasing stocks of material in circulation and increasing modification of the environment by the community There are many exceptions to these trends in particular successions See ECOLOGICAL SUCCESSION

Climax The stable ecosystem which ends the succession is termed climax Climax communities are characterized by self maintenance and a considerable degree of permanence when free from external disturbance That is the populations in the climax stage reproduce and maintain themselves as they do not during the successional stages The climax thus represents a steady state of energy flow materials circulation and population reproduction It may also represent a maximum of sustained productivity for the ecosystem The climax has a self stabilizing character and tends to return to normal after disturbance of its equilibria For instance a heavy phosphate fertilization of a lake produces only a temporary increase in productivity, and phosphate in the water rapidly declines back to the original level Thus a temporary over population of a given species is counteracted by increased mortality from predation disease or other factors until the population returns to its normal level It appears that these stability mechanisms in general depend not so much on any single factor or interaction as on the reestablishment of a steady state balance in relation to various and complexly interrelated factors Relative stability of the community appears to be greater in such highly developed communities of high species diversity as tropical rainforests It is lower in less highly developed communities of lower species diversity and less stable environment as for example, the arctic tundra Characteristics of a climax community are determined in relation to its whole environmental complex A pattern of climax communities may correspond to the pattern of environments in any large area Within a given area, however many environments may be similarly affected by the general climate of the area Many of the successions of a given area consequently converge toward similar climax communities which occupy the largest part of the landscape surface In the low mountains referred to previously the oak forests would occupy the greatest part of the mountain slopes give the vegetation its predominant character and express the climate Such a community may be termed a climatic climax or prevailing climax See CLIMAX COMMUNITY

[R.H W]

Bibliography W C. Allee et al, *Principles of Animal Ecology*, 1949, G L. Clarke *Elements of Ecology* 1954 L. R Dice *Natural Communities*, 1952 F P Odum *Fundamentals of Ecology* 1953.

Eczema

A nonspecific term used to denote any skin disorder characterized by redness thickening oozing from blisters or papules and occasional formation of fissures and crusts Various authorities differ as to the specific skin lesions which are considered under this catch all heading

Infantile eczema may be a form of contact dermatitis atopic or neurodermatitis or combinations of these with others of known or unknown cause

Contact dermatitis occurs in many individual as the result of either chronic irritation or sensitization by some external substance Poison ivy industrial dermatoses and cosmetic sensitivities are examples

Neurodermatitis may be either localized or general in its manifestations An initial itching or pruritis is followed by a thickening of the skin due to repeated scratching There is often a prior history of psychic disturbance or a family history of allergy such as hay fever or asthma

Stasis dermatitis is primarily the result of poor circulation and is found principally on the legs and ankles Older patients or those with an occupation that requires prolonged periods of standing appear to be especially susceptible

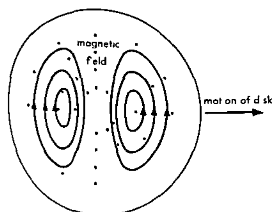
The senile eczema seen in elderly people is thought to be caused by various factors notably dryness of the skin soap sensitivity, poor hygiene or diet or a preceding neurodermatitis

Few eczemas represent pure forms if they have been present for a while If the inciting agent can be found early the prognosis is good, otherwise the usual course is toward a prolonged recurrent or *chronic* disorder complicated by secondary reaction to treatment infection and so on Many of these are well tolerated if the principal symptom that of severe itching is controlled *See* SKIN DISORDERS [E C ST]

Eddy current

An electric current induced within the body of a conductor when that conductor either moves through a *nonuniform magnetic field* or is in a region where there is a change in magnetic flux It is sometimes called Foucault current Although eddy currents can be induced in any electrical conductor the effect is most pronounced in solid metallic conductors Eddy currents are utilized in induction heating and to damp out oscillations in various devices

Causes If a solid conductor is moving through a nonuniform magnetic field emfs are set up that are greater in that part of the conductor that is moving through the strong part of the field than in the part moving through the weaker part of the field Therefore at any one time in the motion there are many closed paths within the body of the conductor in which the net emf is not zero There are thus induced circulatory currents that are called eddy currents (see illustration) In accordance with Lenz s



Eddy currents in a disk moving through a nonuniform magnetic field

law these eddy currents circulate in such a manner as to oppose the motion of the conductor through the magnetic field (*see* LENZ S LAW) The motion is damped by the opposing force For example if a sheet of aluminum is dropped between the poles of an electromagnet it does not fall freely but is retarded by the force due to the eddy currents set up in the sheet If an aluminum plate oscillates between the poles of the electromagnet it will be stopped quickly when the switch is closed and the field set up The energy of motion of the aluminum plate is converted into heat energy in the plate

Eddy currents are also set up within the body of a material when it is in a region in which the magnetic flux is changing rapidly as in the core of a transformer As the alternating current changes rapidly there is also an alternating flux that induces an emf in the secondary coil and at the same time induces emfs in the iron core The emfs in the core cause eddy currents that are undesirable because of the heat developed in the core (which results in high energy losses) and because of an undesirable rise in temperature of the core Another undesirable effect is the magnetic flux set up by the eddy currents This flux is always in such a direction as to oppose the change that caused it and thus it produces a demagnetizing effect in the core The flux never reaches as high a value in the core as it would if there were no eddy currents

Laminations Induced emfs are always present in conductors that move in magnetic fields or are present in fields that are changing However it is possible to reduce the eddy currents caused by these emfs by laminating the conductor that is by building the conductor of many thin sheets that are insulated from each other rather than making it of a single solid piece In an iron core the thin iron sheets are insulated by oxides on the surface or by thin coats of varnish The laminations do not reduce the induced emfs but if they are properly oriented to cut across the paths of the eddy currents they confine the currents largely to single laminae where the paths are long making higher resistance, the resulting net emf in the possible closed path is small Bundles of iron wires or powdered iron formed into a core by high pressure are also used to break up the current paths and reduce the eddy currents *See* CORE LOSS [K V M]

Bibliography S S Attwood *Electric and Magnetic Fields* 3d ed 1949

Edema

An abnormal accumulation of fluid in the cells tissue spaces or cavities of the body also known as dropsy An excess of fluid in the pleural spaces is referred to as hydrothorax in the pericardial sac as hydropericardium and in the peritoneal cavity as ascites Anasarca is a generalized subcutaneous edema

There are three main factors in the formation of generalized edema and a fourth which plays an important role in the formation of local edema They are (1) permeability of the capillary wall (2) colloid osmotic pressure of the plasma proteins (3) hydrostatic pressure in the capillaries and (4) lymphatic obstruction

Permeability of capillary wall Normally the capillary walls are freely permeable to water salts and dissolved gases but are almost impermeable to proteins When the vessel wall is injured by toxins anoxia or paralytic dilatation the capillary endothelium becomes permeable to proteins With

a diffusion of protein into the tissues the plasma osmotic pressure is lowered and the osmotic pressure of the tissue is increased Under these circumstances fluid collects in the tissue spaces

Such a condition plays an important role in inflammatory edema It is a factor in the edema of severe infections metabolic intoxications asphyxia anaphylactic reactions secondary shock and acute nephritis It also contributes to the edematous conditions when there is a fall in the level of plasma proteins

Osmotic pressure of plasma proteins A fall in plasma proteins tends to decrease the forces tending to reabsorb and hold fluid in the vascular compartment Albumin is the protein of *greatest* importance in this regard When the plasma protein level drops below 3 g/100 ml the colloid osmotic pressure is no longer sufficient to maintain a balance with the hydrostatic pressure of the blood which tends to drive fluid out into the tissue spaces Hence more fluid goes out into the tissue spaces and remains there until a new equilibrium is reached

This form of edema is seen in association with prolonged malnutrition (nutritional edema) as during a famine or with chronic nutritional or *metabolic defects* With a marked loss of albumin *in the* urine as in chronic Bright's disease (wet *nephritis* nephrosis) there follows a lowering of *the* plasma albumin fraction and the development *of nephrotic* edema

In nephrotic edema the protein content and [illegible]

level with a relatively greater drop in the albumin *fraction As the plasma osmotic* pressure drops water passes *into* the tissues and with it crystalloids *These substances especially* sodium chloride are retained *in the tissues Therefore* as water is taken in it passes rapidly into the tissues *and* not into the urine

Increased capillary hydrostatic pressure Under normal conditions *the hydrostatic pressure* in the arterial end of the capillary *is sufficient to* overcome the plasma osmotic *pressure and drives* fluid out of the capillary into the tissue *spaces* During passage through *the capillary the pressure* drops to a level low enough *to allow the osmotic* pressure of the proteins to draw *fluid back* into the vascular compartment (*see* CIRCULATION) An increase in hydrostatic *pressure at the* venous end of the capillary will upset *the balance resulting* in a decreased absorption of *tissue fluid by the* osmotic pressure of the plasma proteins *Under these* circumstances an increased *amount of fluid will* be returned via the lymphatics but *as the condition* progresses edema will *develop Such a situation* can follow venous congestion *of long duration*

Cardiac edema following *the generalized* venous congestion of cardiac failure *is the* most common form of this type of edema *The fluid which* collects in the tissue spaces is affected by changes



Photomicrograph of the lung with pulmonary edema (high power view) Compare edema fluid (E) with the unstained air space (A) alveolar wall (W) separates edematous areas

in position being more marked in the dependent portions of the body Fluid also collects in the serous cavities The lymphatics empty into the venous system and therefore a rise in pressure in the venous system results in an increased pressure within the lymphatics which also contributes to the edema formation As the capillary walls are distended they become more permeable. In addition a state of chronic hypoxia may exist causing a further insult to the capillary endothelium with a loss of protein into the tissue spaces

Other factors seem to play a role in cardiac edema With a decreased cardiac output there is a reduced renal blood flow and glomerular filtration rate with a consequent reduced excretion of *salt* and water This may be responsible for an increased volume of extracellular fluid and *plasma* which in turn is followed by a rise in venous *pressure*

Pulmonary edema edema within the *lung* is usually a form of cardiac edema but may be *secondary* to other factors such as inflammation *For a* discussion of pulmonary edema *see* LUNG DISORDERS

Another condition resulting from an *increased* hydrostatic pressure in the capillaries is *postural* edema This occurs when an individual *has been* standing motionless for a long period of *time the* fluid collects in the subcutaneous tissues *of the* feet and ankles

Cirrhosis of the liver causes an *impediment to* the flow of blood through the portal *circulation* There is a consequent rise in venous *pressure and* ascites forms For a discussion of this *condition see* LIVER DISORDERS

Lymphatic obstruction A portion of *the intercellular* tissue fluids returns to the circulation *via* the lymphatics Obstruction to this *channel will* contribute to a local edema Filarial *infection is* one cause of lymphatic obstruction *particularly* in the tropics Lymphatic channels may *be de*

stroyed or obstructed by surgical procedures results in a localized edema Milroy s disease is a chronic hereditary edema thought to be due to lymphatic obstruction *See* FILARIASIS

[R A V]

Bibliography W A D Anderson (ed) *Pathology* 3d ed 1957 W Boyd *Pathology for the Physician* 6th ed 1958 W Boyd *A Text book of Pathology* 6th ed 1953 S Wright *Applied Physiology* 9th ed 1952

Edentata

An order of mammals that includes the anteaters tree sloths and armadillos among living mammals and also the extinct paleanodonts ground sloths and glyptodonts The edentates are exclusively a New World group that originated in North America in the early Tertiary and entered South America before the land connection with North America was broken In South America they underwent a great adaptive radiation during the Tertiary and still form one of the dominant elements of the fauna of that continent

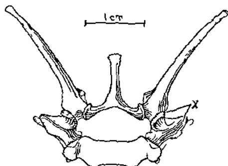
The edentates form a natural but highly diversified group so difficult to characterize that early classifiers included forms such as the aardvark and pangolins which are now known to be unrelated to true edentates The most consistent features of the edentates are the absence of enamel on the teeth and the presence of extra zygapophyses on the lumbar vertebrae The anteaters are the only truly edentulo is edentates

The order is divided into two suborders The Palaeanodonta is a small group of primitive North American edentates that became extinct in the Oligocene They probably represent an unsuccessful side branch of the early edentate stock The Xenarthra, which includes all the remaining edentates is in turn divided into two infraorders The Pilosa include the ground sloths tree sloths and anteaters The sloths are herbivorous mostly leaf eating creatures The anteaters are highly specialized for a diet of termites and ants In the Cingulata the body is encased in a heavy armor formed of bony plates covered with horny scutes The cingulates include the armadillos and the extinct glyptodonts *See* EDENTATA FOSSILS EUTHERIA [D D D]

Edentata fossils

Grouped in this order are many strange fossil animals found exclusively in the Western Hemisphere Edentate meaning toothless applies only to the infraorder Vermilingua (worm tongue South American anteaters) Other edentates have simple prismatic teeth in which the enamel is lost (except in a few early forms and in the embryos of living armadillos) There was little dental specialization into incisors canines premolars and molars Some edentates were no larger than rats others (megatheres ground sloths) rivaled elephants in size All were clawed or secondarily hoofed and were quadrupedal Some probably lived in trees others were burrowers Comparisons with living forms indicate that many were probably insectivorous while others fed more upon vegetation

In the suborder Xenarthra (strange joint) which includes all edentates except the early Tertiary North American Palaeanodonta the posterior vertebrae of the back have extra articular facets These are found in no other mammalian order Xenarthran vertebrae also tend to be fused in the shoulder and hip regions with the extreme condition occurring in glyptodonts where the two members of each set of vertebrae are fused Most peculiar among the many xenarthran specializations, however were the reduction of hairy covering and development of bony and horny armor This armor varied from the condition of scattered bony ossicles in the skin (Mylodontidae ground sloths) to the mobile carapace of armadillos (Dasypodidae with imbricating bands of scutes in the thoracic region)



Posterior view of a lumbar vertebra of a late Miocene armadillo from Colombia South America X xenarthral articulation facets

and the semirigid dorsal shield of the Glyptodontidae Xenarthrans include Vermilingua (mid Miocene into Recent) Gravigrada (ground and tree sloths late Eocene into Recent) and Cingulata (armadillos and glyptodonts late Paleocene into Recent)

Edentates probably originated in North America in an early to mid Paleocene stock of small generalized clawed quadrupedal mammals that would be included at present in the order Insectivora *Palaeanodon* (North American late Paleocene and early Eocene) qualifies in most details as a prototype edentate Specialized palaeanodonts survived in the United States into Oligocene time Apparently a *Palaeanodon* like stock arrived in South America in earliest Cenozoic time before that continent became completely isolated by sea ways there to give rise to all of the xenarthrans South America is the homeland of xenarthrans their earliest and principal records are there It was not until late Cenozoic time that such groups as ground sloths glyptodonts and armadillos reinvaded North America A unique group of Pleistocene sloths in Cuba and Puerto Rico (*Megalocnus Acratocnus* and others) suggests a mid Cenozoic dispersal from South America to the Antilles *See* INSECTIVORA FOSSILS

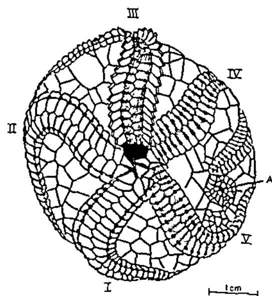
Eastern Hemisphere "edentates" the pangolins are now placed in a separate order the Pholidota and appear to have a race history that is distinct from that of the true edentates *See* PHOLIDOTA FOSSILS [D F S]

Edison battery

A storage battery composed of cells having nickel and iron in an alkaline solution also called a nickel iron alkaline battery The active material on the negative plates is iron and that on the positive plates is nickel oxide The electrolyte is potassium hydroxide During discharge the nickel oxide is reduced to a nickelous hydroxide and the iron is oxidized to ferrous hydrate During charge the reverse process takes place The Edison battery is lighter in weight than the more common lead storage battery and has longer life because there is no chemical deterioration during charge and discharge Edison storage batteries are used chiefly in electric industrial trucks and tractors mine locomotives railroad signal and lighting service telephone service isolated airway beacons and emergency power for lighting police and fire alarm systems *See* STORAGE BATTERY [J MR]

Edrioasteroidea

A class of extinct Pelmatozoa which arose early in Cambrian times and died out before the end of the Carboniferous Period They differ from all other Pelmatozoa in having the ambulacral grooves bordered by tube feet which emerged through pores in or between the ambulacral plates In this feature they resemble Eleutherozoa and they may be closely related to them They differ from Eleutherozoa in having the mouth and anus on the upper side of the theca Hinged ambulacral plates bordered the groove on either side and could be erected over it to convert it into a tube The theca shows strong pentamerous symmetry and is more or less discoidal It bears no stem The animals were in some cases cemented to the substrate by the lower surface in other cases they were free and rested upon the sea bed Brachioles are lacking About 30 genera have been described and grouped into 7 families *See* ELEUTHEROZOA, PELMATOZOA [H.B F]



An edrioasteroid (*Edrioaster*) seen from above The mouth is at the center and the 5 ambulacra are numbered I to V In ambulacra I and II the covering ambulacral plates are in position in III they are shown hinged in the lateral position to expose the groove in IV and V the hinged plates are removed leaving the adambulacral plates exposed and showing the pores A anus (*Modified from Bather 1900*)

Eel

Any of a number of fishes comprising the order Anguilliformes characterized by long slender bodies small scales or none at all usually no pelvic fins continuous dorsal caudal and anal fins and small gill openings There are at least three families of eels mostly marine



The common eel *Anguilla rostrata* length to 5 ft (*From E L Palmer, Fieldbook of Natural History McGraw Hill 1949*)

The fresh water eels of America and Europe are similar *Anguilla rostrata* the American eel occurs in most of the streams of the Atlantic drainage from Newfoundland to Central America and in the West Indies The females swim up stream great distances and are found in virtually all of the streams of the Mississippi drainage The males seldom enter fresh water When mature the eels migrate to an area in the Atlantic near the Bermuda Islands where they spawn The small transparent larvae called leptocephali appear in great numbers here Gradually they make their way to the coast where they transform into small eels called elvers when about 1 year old Most American eels are mature when about 3 ft long but 5 ft specimens are known Similarly the European eel migrates to the same area to spawn and requires 2 years to mature Eels are cultured extensively in special ponds in Europe where they are held in much higher esteem for food than in America

The moray eels are all marine and mostly tropical and are found most commonly on coral reefs They have thick leathery skins are often brightly colored and have strong sharp teeth Some of them are large vicious fishes which are considered dan

gerous to divers and fishermen who invade their habitat

The conger eels are scaleless marine eels usually found in warm waters although one species ranges northward to Cape Cod

The electric eel *Electrophorus electricus* is not a true eel but belongs to the order Cypriniformes with the suckers, minnows and catfishes It occurs in the drainage basins of the Amazon and Orinoco rivers where it reaches a length of 8 ft This fish is said to produce the most powerful electric shock of any of the electric fishes and is credited with killing horses as well as humans when they invade its home waters *See* ANGUILLIFORMES CYPRINIFORMES [J D B]

Effective dose 50

This term is used chiefly to characterize the potency of a drug by the amount required to produce a response in 50% of the subjects to whom the drug is given The term is also known as ED_{50} or median effective dose At one time it was usual to try to measure the effect of a drug by noting the amount which was just sufficient to produce a particular response but when it was realized that this amount varied greatly from subject to subject attention was turned to measuring the effect on a group of subjects Suppose for example that a drug is being used to relieve a certain type of pain then the median effective dose is of such a size that it controls the pain in 50% of the sufferers and is insufficient to control it in the remaining 50%

The term median effective dose is most commonly applied in connection with drugs but it may be used when various other sources of stimuli for example x rays are under consideration The response must be of the kind known as quantal or all-or nothing where the investigator is simply able to report that the response either was or was not elicited for example convulsions did or did not occur hemorrhage was or was not produced pregnancy did or did not ensue the animals did or did not survive

The median effective dose is not a well defined quantity until the test animal the end response and such factors as the route of injection of a drug and the state of nutrition of the animal are specified In the work of an investigator who uses the appropriate controlled conditions however the ED_{50} is a reproducible measurement

Determination of the ED_{50} of a drug requires the administration of at least two separate amounts of the drug each one given to several subjects Suppose that an animal physiologist wishes to measure the ED_{50} of a hormone estrone that causes estrus or heat He has specified that he will use spayed female rats as experimental animals and has also laid down various conditions that he regards as important in controlling the results of his investigation the age of the animals perhaps and the route by which the hormone will be administered It is unlikely that a reliable measurement could be made with fewer than 30 animals Suppose that a total of 60 were used 20 at each of 3 doses with the following results

Dose international units	Animals showing estrus %
2	20
3	45
4	80

Inspection of the results suggests that a dose somewhat greater than 3 units will cause estrus in 50% of the animals Methods based on the theory of probability can be used to give a more objective analysis of the data and to provide a measure of the error of the estimate but it is sufficient to note for the present purpose that the ED_{50} is approximately 3 It is clear that in measuring the ED_{50} some choices of the dosage to be used will work out better than others Doses giving a response in a very small or very large percentage of the subjects contribute little information to the measurement of ED_{50} The ideal is to have doses on either side of the ED_{50} and the closer they are to ED_{50} the better However in practice a pharmacologist sometimes determines an ED_{50} as one part of a larger experiment to investigate the dose response curve of a drug or to compare the potency of two preparations of a drug In such a case though he would still want to have doses on each side of the ED_{50} he would not want them close to the ED_{50} More information about the total curve would be given if the low dose had an effect in say 25% of the animals and the high dose in about 75% *See* BIOASSAY [C W H]

Effector systems

Those organ systems of the animal body which mediate overt behavior Injury to an effector system leads to loss or to subnormal execution of behavior patterns mediated by the system conditions termed paralysis and paresis respectively

Overt behavior consists of either movement or secretion Movement results from contraction of muscle Secretion is a function of glands Neither muscular contraction nor glandular secretion is autonomous but instead is regulated by an activating mechanism which may be either neural or humoral In neurally activated systems the effector organ whether muscle or gland is supplied by nerve fibers originating from cell bodies situated in the central nervous system or in peripherally located aggregates of nerve cell bodies known as ganglia The nerve fibers make intimate contact with but are not protoplasmically continuous with the cells of the effector organ Activation of the effector organ occurs when the nerve cell body is excited and generates a nerve impulse an electrochemical alteration which is conducted along the

this alteration in turn leads to either contraction or secretion Such systems comprised of a muscle or a gland along with their regulating nerves are termed neuromuscular or neuroglandular effector systems respectively Examples are the skeletal muscles together with their motor nerve supplies and the medullae of the adrenal glands along with their innervation (splanchnic nerves) In many neurally regulated effector systems (for example skeletal muscle and adrenal medulla) function is totally dependent on intact innervation and denervation leads to functional paralysis In other organs (for example the salivary glands) denervation causes only temporary paralysis When recovery occurs the gland may oversecrete continuously (paralytic secretion) apparently because the denervated gland cells become unusually sensitive to certain blood borne chemical agents (denervation hypersensitivity) See ADRENAL GLAND BIOPOTENTIALS AND ELECTROPHYSIOLOGY

In other effector systems (humeromuscular humeroglandular) the activating agent is normally a blood borne chemical substance produced in an organ distant from the effector organ Uterine smooth muscle is uninfluenced by the uterine nerve activity but contracts vigorously when the blood contains pitocin a chemical substance elaborated by the posterior lobe of the hypophysis sensitivity to pitocin increases progressively during pregnancy Similarly secretion of pancreatic juice is independent of pancreatic innervation the regulating agents are blood borne substances (pancreozymin and secretin) produced by cells in the wall of the small intestine Generally in such humorally regulated effector systems activation is more delayed and more prolonged than in neurally regulated systems See HYPOPHYSIS UTERUS

Finally some effector systems are hybrid in the sense that both nerves and humors regulate their functions The smooth muscle of arterioles contracts in response to either nerve stimulation or epinephrine a substance secreted into the blood stream by the adrenal medulla Secretion of hydrochloric acid by the gastric mucosa is increased by activation of the vagus nerve or by the presence in the blood of histamine a substance found in many tissues of the body Effector systems with both neural and humoral regulation are never completely paralyzed by denervation but may be deficient in reaction patterns when the quick integrated activation provided by neural regulation is essential For example following extensive vascular denervation humoral agents may maintain sufficient arteriolar constriction to sustain the blood pressure in static situations However the normal capacity to increase arteriolar constriction to offset the gravitational effects of rising from the prone to standing positions is permanently lost with the

TEM PSYCHOLOGY PHYSIOLOGICAL AND EXPERIMENTAL

[HDP, TCR]

Efficiency

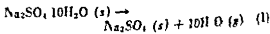
The ratio, expressed as a percentage of the power output to the power input When only mechanical efficiency is concerned the difference or loss between the input and output power is due to friction and is dissipated in the form of heat (see SIMPLE MACHINE)

[RMFH]

Efflorescence

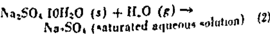
The spontaneous loss of water (as vapor) from hydrated crystalline solids The thermodynamic requirement for efflorescence is that the partial pressure of water vapor at the surface of the solid (its dissociation pressure) exceed the partial pressure of water vapor in the air

A typical efflorescent substance is Glauber's salt, $Na_2SO_4 \cdot 10H_2O$ At 25°C, the dissociation pressure for the process in Eq (1)



$$Na_2SO_4 \cdot 10H_2O\ (s) \rightarrow Na_2SO_4\ (s) + 10H_2O\ (g) \quad (1)$$

is 19 4 mm Hg 81% of the saturation vapor pressure of pure water at this temperature In a sufficiently humid atmosphere Glauber's salt also can deliquesce by the process shown in Eq (2)



$$Na_2SO_4 \cdot 10H_2O\ (s) + H_2O\ (g) \rightarrow Na_2SO_4\ \text{(saturated aqueous solution)} \quad (2)$$

The vapor pressure of the saturated solution is 21 9 mm Hg 92% of the vapor pressure of pure water Thus Glauber's salt at 25°C is stable in atmospheres having relative humidities of 81–92%, below 81% it effloresces, above 92% it deliquesces

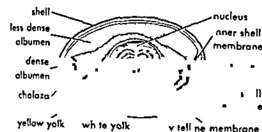
The spontaneous loss of water normally requires that the crystal structure be rearranged and consequently efflorescent salts usually go to microcrystalline powders when they lose their water of hydration See DELIQUESCENCE, EQUILIBRIUM PHASE VAPOR PRESSURE

[RLS]

Egg (fowl)

A single large living female sex cell enclosed in a porous, calcareous shell through which gases may pass See CELL (BIOLOGICAL) Although they vary in size shape and color, the eggs of chickens ducks geese and turkeys are essentially the same in structure and content Inward from the shell are the outer and inner shell membranes which are also permeable to gases The membranes are constructed to prevent rapid evaporation of moisture from the egg but to allow free entry of oxygen which is necessary for life (see RESPIRATION) Air begins to penetrate the shell soon after the egg is laid and it tends to accumulate in a space between the two membranes at the large end of the egg

The inner shell membrane surrounds a mass of fluid albumin which in turn encloses a body of dense albumin, these two types of protoplasm constitute the so called egg white (see PROTOPLASM) The central part of the egg is occupied by the yolk which contains the vital egg nucleus and its associated parts (see CELL NUCLEUS) The



Egg of a b rd (After Sch mkewitsch from L P Sayles ed B ology of the Vertebrates 3d ed Macm llan 1949)

yolk consists of alternating layers of yellow and white yolk The yolk enclosed by the vitelline membrane is held in place by the chalaza which is anchored at each end of the egg and prevents un due mechanical disturbance

When a sperm nucleus fuses with the egg nu cleus, the process is called fertilization (*see* FERTILIZATION) Within a few hours or less the fertilized egg begins a series of cell divisions and d fferentiations which result in the formation of the embryo (*see* CELL DIVISION MITOSIS) The embryo then undergoes further cell modification and eventually develops into the young of the spe cies In time the pressure created by the growth of the embryo causes the shell to rupture and the young hatches As the young emerges from the shell it carries with it a part of the food and water orig nally in the yolk on which it can subsist for a few days After this initial period the young must have access to food and water

For a discussion of the steps involved in the development of various kinds of fertilized eggs into mature embryos *see* EMBRYOLOGY OVUM REPRODUCTION ANIMAL *See also* POULTRY PRODUCTION [J F F]

Egg processing

The commercial procedures used in the collection and d stribution of shell eggs and the further proc essing of shell eggs for industrial uses

Eggs commonly used in the United States are cons dered to be the product of the domesticated ch cken exclusively

Egg statistics Approximately 7 600 000 000 lb of eggs were produced in the United States in 1957 g ving a per capita consumption of 360 eggs per year Of these eggs approximately 90% are con sumed as shell eggs 2 5% are used for hatchery reproduction and 6 5% are converted into frozen or dried eggs In 1957 334 000 000 lb were frozen and 112 000 000 lb were used by dryers to produce the equivalent of 28 000 000 lb of whole egg solids The production of egg solids in 1957 was approxi mately equally divided between whole eggs egg white sol ds and egg yolk sol ds

Industry use constituents nutrition Frozen eggs and egg solids find their primary uses in bak ery goods doughnuts noodles mayonnaise con fectionery items and prepared mixes Eggs are used in bakery products for their binding action leavening action emulsifying action flavor color and nutritive value Several quality attributes are of utmost importance to the user of egg products for example chemical composition nutritional value microbiological properties organoleptic proper ties and functional performance Chemically liq

cose) A great many other minor constituents are also present

Eggs are an excellent source of nutritional ele ments with all vitamins present in substantial quantities except vitamin C. The lipids of the egg yolk are composed mainly of highly unsaturated fatty acids containing a large proportion of the essential fatty acid oleic acid Egg protein is rec ognized as a biological standard in nutritional re search *See* NUTRITION

Commercial egg production Large quantities of eggs are produced on farms in the Middle West but recent production shifts have resulted in a large proportion of shell egg production near heavily populated urban areas These shifts to commercial large scale production result in the preparation of egg products in the Middle West from surplus farm eggs

Eggs are graded for quality by candling then classified as to size placed in cartons and dis tributed under refrigeration to retail outlets Grad ing and size classification are carried out under U S Department of Agriculture or state standards in most plants

Eggs for the products industry These are bro ken either by hand or by breaking machines with the whites and yolks being collected separately After thorough mixing and standardization for solids content color or both eggs for frozen con sumption are placed in 30 lb tin containers and sharp frozen at −20 to −40°F in blast freezers Distribution of these frozen eggs follows normal wholesale food channels Egg liquid to be dried is either dried on the premises or hauled by in sulated tank truck to centrally located egg drying plants

Egg white drying Prior to drying egg white liq uid undergoes extensive processing to insure the retention of the functional and organoleptic prop erties in the finished product A small amount of free glucose must be removed to prevent the Maillard reaction Glucose removal is accom plished by the use of bacteria or yeast capable of utilizing glucose or by enzyme systems capable of converting the glucose to a nonreactive material such as gluconic acid Following appropriate ad justment of pH with lactic citric or other ap proved food acids the liquid egg white is spray dried in conventional dryers of the Rogers or simi lar type

Egg white solids are normally packed in 150 lb fiber drums or approximately 25 lb polyethylene

lined boxes for shipment to food manufacturers. Egg white is also air dried in so called pan or tunnel systems in which the egg is poured in a thin layer on trays. After appropriate liquid treatment as described above these trays are subjected to temperatures up to 130°F until a pseud crystalline product is obtained. These crystals or flakes are used as such or ground into a fine powder for greater solubility. This type of product finds its primary use in the confectionery industry.

Egg yolk drying. Liquid egg yolk from the breaking points is collected at drying plants in a similar manner where it may be pasteurized at 143°F for 3 min and spray dried directly or it may be subjected to a stabilizing treatment to remove the free glucose. The glucose oxidase enzyme system is used for this process almost exclusively. Egg yolk can be dried in many different kinds of spray drying equipment for example the Rogers Gray Jensen and Moajonnier systems.

Whole egg drying. Whole egg solids are manufactured in a similar manner as are many blends of whole egg yolk and added ingredients to meet specific functional requirements. Common additives are carbohydrate products which may be either sucrose or corn syrup solids derivatives. These carbohydrates when added prior to spray drying insure that the original foaming and emulsifying abilities of the liquid egg are unimpaired. See FOOD ENGINEERING. [K H FO]

Eggplant

A warm season vegetable (*Solanum melongena*) of Asiatic origin belonging to the plant order Tubiflorales. Eggplant is grown for its egg shaped fleshy



Eggplant *Solanum melongena*

fruit and is eaten as a cooked vegetable. Cultural practices are similar to those used for tomatoes and peppers however eggplant is more sensitive to low temper [illegible] d varie [illegible] fruit of [illegible] yellow and green are used chiefly for ornamental purposes. Harvesting generally begins 70–80 days after planting. Florida and New Jersey are important producing states. The total annual farm value in the United States is approximately $2,200,000. See PEPPER, TOMATO, TUBIFLORALES, VEGETABLE GROWING. [H J C]

Eigenfunction

In utmost generality, a (not identically zero) solution ψ to an equation possessing solutions ψ only for special values of a parameter λ, the eigenvalues. It is also termed characteristic function. Usually the eigenvalue equation takes the form $A\psi = \lambda B\psi$ with A and B being linear operators, and the eigenvalues determined by the requirement that solutions ψ satisfy imposed boundary conditions.

In quantum mechanics where B typically is unity and A is the Hamiltonian the equation is Schrödinger's, ψ is the wave function λ is the energy, and the boundary conditions guarantee there are no solutions when λ is a complex number. Similar equations also occur in the theory of vibrating systems and in heat conduction and diffusion problems in classical physics. See EIGENVALUE, QUANTUM THEORY, NONRELATIVISTIC. [E.G.]

Eigenvalue

A possibly complex number λ for which the eigenvalue equation usually of form $A\psi = \lambda B\psi$ has a (not identically zero) solution ψ satisfying the boundary conditions. It is also termed characteristic value. An eigenvalue for which more than one independent solution exists is degenerate. For an example in the interval $0 \leqq x \leqq \pi$ $\psi(x,\lambda) = \sin \sqrt{\lambda}x$ or $\cos \sqrt{\lambda}x$ satisfies $-d^2\psi/dx^2 = \lambda\psi$ for any λ real or complex. The boundary conditions $\psi = 0$ at $x = 0$ and $x = \pi$ restrict the eigenvalues to $\lambda = 1, 4, 9, \ldots, n^2, \ldots$ with but one eigenfunction $\psi(x, n^2) = \sin nx$ for each eigenvalue n^2; thus these eigenvalues are not degenerate. See DEGENERACY (QUANTUM STATES), EIGENFUNCTION, QUANTUM THEORY, NONRELATIVISTIC. [E G]

Eimeriida

An order of protozoans in the class *Coccidia*. The complete life cycle including schizogony, gamete formation and fertilization is passed in one host. These parasites infest man as well as economically important animals such as sheep, cattle, poultry and fur bearing animals.

Life cycle. In the life cycle (see SPOROZOA) the walled zygotes now called oocysts usually leave the host in the excrement where sporogony then ensues. Temperature, humidity and oxygen concentration are important factors regulating sporogony. At its completion the sporulated oocysts contain the infective sporozoites which may be transmitted to the new host as chance contaminants of its food or water or by its licking the hide of an animal soiled with filth containing them.

The genus *Eimeria* is characterized by sporulated oocysts with four spores each with two sporozoites and *Isospora* by two spores each having four sporo-

zoites Most species of *Eimeria* and *Isospora* parasitize epithelial cells of a particular region of the digestive tract below the stomach notable exceptions being *Eimeria stiedai* of the biliary epithelium of the rabbit liver and *Eimeria truncata* of the goose kidney The preference shown for particular areas or organs is called organ specificity Each species of coccidium also shows a highly exclusive preference for certain species of hosts which is called host specificity On the other hand one host species may serve for more than one species of coccidium that is the chicken harbors at least 8 species of *Eimeria* the turkey 7 the rabbit about 12 and the cow about 11 Dogs and cats can harbor at least two species of *Eimeria* and three of *Isospora*

Importance Certain species particularly in heavy infections can do immeasurable harm to their hosts while other species are well tolerated *Eimeria tenella* with affinities for the cecal mucosa of chickens can cause severe or fatal coccidiosis in chicks characterized by hemorrhage into the ceca and bloody droppings *Eimeria stiedai* of the rabbit s liver can so severely derange the function ing of that organ that death or permanent unthriftiness may result Cattle and sheep each harbor several different species which under certain conditions cause severe losses On the other hand each of these hosts harbors a number of species of slight consequence to their health Man is infrequently the host for *Isospora belli* and still more infrequently for *Isospora hominis* Comparatively heavy infections of either parasite can cause gastronomic disturbances See COCCIDIA, see also ADELEIDA

[E R BE]

Einstein shift

A shift towards longer wavelengths of spectral lines emitted by atoms in strong gravitational fields One of three famous predictions of the general theory of relativity this shift results from the slowing down of all periodic processes in a gravitational field The amount of the shift is proportional to the difference in gravitational potential between the source and the receiver For starlight received at the Earth the shift is proportional to the mass of the star divided by its radius In the solar spectrum the shift amounts to about 0 01 angstrom (A) at a wavelength of 5000 A In the spectra of white dwarfs whose ratio of mass to radius is about thirty times that of the Sun the shift is about 0 3 A which can easily be measured if it can be separated from the Doppler effect This was done by W S Adams for the companion of Sirius a white dwarf whose true velocity relative to the Earth can be deduced from the observed Doppler effect in the spectrum of Sirius The measured shift agreed with the prediction based on Albert Einstein s theory and on independent determinations of the mass and radius of Sirius B Attempts to demonstrate the Einstein shift in the solar spectrum have thus far proved inconclusive because it is difficult to distinguish the Einstein shift from so called pressure shifts resulting from perturbations of the emitting atoms by neighboring atoms

Attempts have also been made to deduce stellar masses from measurements of the Einstein shift but the difficulty of allowing properly for the Doppler effect and for pressure shifts renders these determinations very uncertain See RELATIVITY

[D L]

Einsteinium

Element number 99 einsteinium Es is a member of the actinide series of elements It is not found in nature but was discovered and is produced by



artificial nuclear transmutation of lighter elements All isotopes of einsteinium are radioactive decaying with half lives of from a few minutes to about one year The world s supply of the element is an invisible amount—far less than one-millionth of a

[illegible]

mos Argonne and University of California laboratories of the U S Atomic Energy Commission A very high flux of neutrons at the instant of explosion resulted in multiple neutron capture by uranium present in the bomb to give the mass 253 isotope of uranium A subsequent chain of β decays through mass 253 isotopes of intervening elements finally gave Es^{253} Einsteinium was separated from other actinide elements by the ion exchange method and was found to decay by α emission with a half

[illegible]

celerated in a cyclotron See BERKELIUM

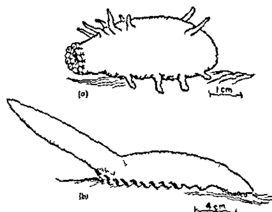
Tracer scale studies of the chemical properties of einsteinium indicate its existence in the 3+ oxidation state characteristic of actinide elements As do other tripositive actinides einsteinium coprecipitates with lanthanum fluoride or hydroxide and forms a chelate complex with thenoyltrifluoroacetone (TTA) which can be extracted into benzene For the separation and identification of einsteinium in the presence of other actinides ion exchange chromatography has been of considerable value Actinides are desorbed from cation exchange resin at different rates when a complexing agent is

passed through the column of resin As a result the elements leave the column in distinct fractions and can be identified by their characteristic radioactivity See ION EXCHANGE NUCLEAR REACTION TRANSURANIUM ELEMENTS [S G T]

Bibliography J J Katz and G T Seaborg *The Chemistry of the Actinide Elements* 1958

Elasipoda

An order of almost exclusively *deep sea Holothuroidea* which have 10-20 *shield shaped tentacles* but lack respiratory *trees and pharyngeal retractor* muscles *Tube feet* are usually present



Deep sea Elasipoda (a) Elpidia (b) Psychropotes

Elasipoda occur in all the oceans but *are especially* significant elements in the faunas of *deep trenches* ranging down to 5000 fathoms *probably because* they more than other animals can extract *nutriment* from mud in an environment *almost* devoid of other food The bottom dwelling *forms* such as *Elpidia* have an ovoid body *or as in Psychropotes* the body may be rather flattened and furnished with a finlike posterodorsal tail *which* is probably used in plowing through mud *Pelagothuria* is a pelagic form which swims *with the aid of a* web supported on long papillae around the mouth *The order* embraces five families *See* HOLOTHUROIDEA [H B F]

Elasmobranchii

The larger and more important *group* of the two subclasses of cartilaginous *fishes* (Chondrichthyes) including the sharks and rays The elasmobranchs differ from the other *subclass* the Holocephali or chimaeras in the *possession* of 5–7 pairs of gills and *external gill openings* in the nonerectile dorsal fin or fins *in the* usual presence of denticles or dermal *spines and in* having at least some ribs differentiated *vertebral* centra and a cloaca There is no frontal *clasper* The teeth are numerous and the upper *jaw* is not united with the braincase

The Elasmobranchii includes two *orders* the Squaliformes or sharks and the *Rajiformes* or skates rays sawfishes and guitarfishes There are 500-600 elasmobranch *species See* CHONDRICHTHYES, ELASMOBRANCHII FOSSILS, RAJIFORMES SQUALIFORMES [E M S]

Elasmobranchii fossils

The fossil remains of the subclass Elasmobranchii which includes all fishes of the class Chondrichthyes except the aberrant chimaeras and their presumed bradyodont ancestors and relatives. Elasmobranch remains are known from a great variety of fossil deposits mainly marine from Devonian times on *Knowledge of the group as* fossils is however, much less satisfactory than in the case of most other fishes because of the cartilaginous nature of the skeleton Unless calcified cartilage is rarely preserved in fossil form In consequence most of the information on the history of the group is limited to finds of teeth and spines as the only hard skeletal elements specimens of entire fishes are relatively rare *See* CHONDRICHTHYES *see also* BRADYODONTI CHIMAERAE.

It was formerly believed that the absence of bone in the skeleton was a primitive characteristic and therefore sharks would appear relatively early in the fossil record The converse is the case even members of the progressive bony fish class the Osteichthyes are abundantly present in deposits laid down before the middle of the Devonian *but* there are almost no traces of elasmobranchs until the terminal phases of that period However it is now generally accepted that the absence of bone in elasmobranchs is due to skeletal degeneration *and* hence this late appearance is no longer surprising Presumably elasmobranchs were descended from antecedent forms with well-ossified skeletons Such forms are to be sought amongst the archaic fish class of Placodermi which had come into existence a period earlier None of the typical placoderms among the arthrodires for example is *suitable as* an ancestor, there are however several Devonian marine placoderm types related to the *primitive* arthrodires in which the bony armor was apparently undergoing reduction These if not directly ancestral at least represent stages in shark development *See* OSTEICHTHYES FOSSILS PLACODERMI

Elasmobranch history consists essentially of two chapters There appears to have been a considerable radiation of sharks beginning in the *Late* Devonian and continuing through the Carboniferous followed by reduction and near extinction *of* the group by the end of the Permian *A few* sharks however survived the general reduction *of* all forms of marine life that took place at the end of the Permian and in the Jurassic there *began a* second radiation of elasmobranchs which was fully attained by the late Cretaceous Since that time there appears to have been little change *the definite ... of the* elasmobranch types concerned changed considerably The oldest Devonian forms are almost ex

clusively members of the primitive order, the Cladoselachii In the Carboniferous, it would appear that many of the forms present belonged in the Hybodontoidea a more progressive group with claspers and fin modifications which have caused them to be included among the members of the order Selachii It is difficult, however, to determine to which of the two groups many forms of the late Paleozoic should be assigned in default of complete skeletons for tooth and spine types present in cladoselachians are known to have been present in some of their hybodontoid descendants Many of the latter, however, tended to develop a heterodont dentition with sharp-pointed teeth persisting in the front of the mouth but with cheek teeth flattened and suitable for mollusk feeding It is presumably from such hybodonts that the Bradyodonti arose, and from them the later chimaeras

It was hybodonts of this sort, versatile as to their food supply, which alone survived the difficult times at the end of the Paleozoic They were flourishing again in modest fashion toward the end of the Triassic, only to dwindle in numbers in later periods During the Jurassic, however, there began a radiation from this surviving group which gave rise

paralleled certain of the Paleozoic hybodonts and the contemporary bradyodonts in bottom dwelling and mollusk eating habits

The great majority of elasmobranchs living and fossil, are marine However, some Cretaceous and Eocene skates and rays are found in deposits which suggest that, like certain modern forms they had invaded brackish, if not fresh water, habitats The Pleuracanthodii of the late Paleozoic are the only elasmobranchs which were typically fresh water forms *See* BATOIDEA FOSSILS, CLADOSELACHII, HYBODONTOIDEA, PLEURACANTHODII SELACHII FOSSILS [A S R]

Elastic limit

The greatest stress a material can sustain without causing permanent strain after release of stress Actual materials are not perfectly elastic upon first application of load because they lack homogeneity and possess residual stresses caused by manufacturing treatments *See* STRESS AND STRAIN

[W J KR]

Elasticity

The property whereby a solid material changes its shape and size under the action of opposing forces, but recovers its original configuration when the forces are removed The theory of elasticity deals with the relations between the forces acting on a body and the resulting changes in configuration and is important in many branches of science and technology, for instance in the design of structures in the theory of vibration and sound, and in the study of the forces between atoms in crystal lattices

Elastic constants. The forces acting on a body are expressed as stresses and measured as force per unit area Thus if a bar $ABCD$ of square cross section (Fig 1*a*) is fixed at one end and subjected to a force F uniformly distributed over the other end DC, the stress is $F/(DC)^2$ This stress causes the bar to become longer and thinner and to assume the shape $A'B'C'D'$ The strain is measured by the ratio (change in length)/(original length), that is, by $(B'C' - BC)/(BC)$ According to Hooke's law stress is proportional to strain, and the ratio of stress to strain is therefore a constant, in this case the Young's modulus denoted by E, so that $E = F(BC)/(DC)^2(B'C' - BC)$ *See* HOOKE'S LAW, STRESS AND STRAIN, YOUNG'S MODULUS

Poisson's ratio (σ) is defined as the ratio of lateral strain to longitudinal strain so that $\sigma = BC(DC - D'C')/DC(B'C' - BC)$ The bar of Fig 1*a* is in a state of tension, and the stress is tensile if the force F were reversed in direction the stress would be compressive Stresses of this type are called direct or normal stresses, a second type of stress, known as tangential or shear stress is illustrated in Fig 1*b* In this case, the configuration $ABCD$ becomes $ABC'D'$, with the shear forces F acting in the directions AB and CD The shear strain is measured by the angle θ, and if the body is originally a cube, the shear stress is $F/(DC)^2$ The ratio of stress to strain, $F/(DC)^2\theta$, is the shear or rigidity modulus G, which measures the resistance of the material to change in shape without change in volume

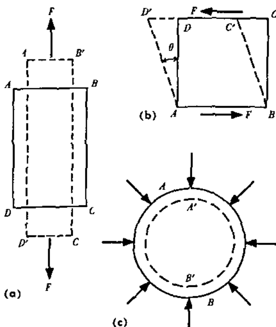


Fig 1 Stresses on a bar (a) Direct or normal stress (b) Tangential or shear stress (c) Change in volume with no change in shape (All deformations exaggerated)

A further elastic constant the bulk modulus k measures the resistance to change in volume without change in shape and is illustrated in Fig 1c The original configuration is represented by the circle AB and under a hydrostatic (uniform) pressure P the circle AB becomes the circle AB The bulk modulus is then $k = Pv/\Delta v$ where $\Delta v/v$ is the volumetric strain The reciprocal of the bulk modulus is the compressibility

Determination of values The elastic constants may be determined directly in the way suggested by their definitions for instance Young s modulus can be determined by measuring the relative extension of a rod or wire subjected to a known tensile stress Less direct methods are however usually more convenient and accurate Prominent among these are the dynamic methods involving frequency of vibration and velocity of sound propagation The elastic constants can be expressed in terms of frequency of (or velocity in) regularly shaped specimens together with the dimensions and density and by measuring these quantities the elastic constants can be found *See* ULTRASONICS

The elastic constants can also be determined from the flexure and torsion of bars As an illustration consider a bar AB (Fig 2) of breadth b (in the x direction) and depth d (in the x direction) supported by forces F at the ends and loaded symmetrically by forces F at points C and D Over the portion CD there is a uniform bending moment $M = Fl_2$ and the theory of bending shows that the portion CD is bent into the arc of a circle such that

$$R = EI/M \qquad (1)$$

where R is the radius of curvature E is Young s modulus and I is the moment of inertia of cross section equal to $bd^3/12$ for a rectangular cross section The longitudinal stress at the lower face of the bar is tensile and at the upper face compressive The middle plane of the bar is free of stress and is the neutral axis The stress at a distance x_2 from the neutral axis is

$$T = Ex_2/R = Mx_2/I \qquad (2)$$

It is thus possible to determine E from Eq (1) by measuring I R and M conversely if E is known the stress may be determined from Eq (2) *See* LOADING TRANSVERSE

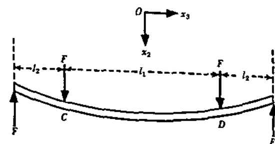


Fig 2 Flexure and torsion in a bar

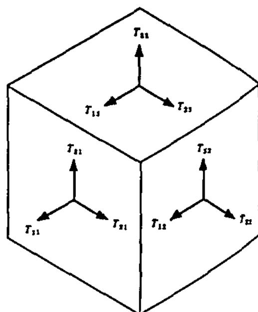


Fig 3 Stress components

Practical limitations In practice stress is only proportional to strain and the strain is only completely recoverable within certain limits called the elastic limits of the material The stress below which the strain is completely recoverable is sometimes called the limit of perfect elasticity and the stress up to which Hooke's law is obeyed is sometimes called the proportional limit or limit of linear elasticity Above the elastic limits the material is subject to time dependent effects and as the stress is further increased the ultimate strength of the material is approached *See* PLASTICITY STRENGTH OF MATERIALS

Theory of elasticity In classical elasticity theory it is assumed that the strains are always small Hooke's law is therefore obeyed the strains are completely recoverable and moreover are superposable so that the strain produced by the joint action of two or more stresses is the sum of the strains produced by them individually

In order to develop the theory it is necessary to specify the stresses and strains more closely Figure 3 shows the stress components T_{ij} (where i j may take the values 1 2 or 3) acting on the faces of a cube parallel to the coordinate axes x_1 x_2 x_3 The first suffix indicates the direction of the stress component and the second the direction of the normal to the plane under consideration Stresses of the type T_{11} are normal stresses and of the type T_{12} shear stresses The conditions for zero rotation of the cube are $T_{12} = T_{21}$ $T_{13} = T_{31}$ $T_{23} = T_{32}$ and there are therefore six independent stress components

In addition to the stresses T_{ij} body forces proportional to volume (for instance forces due to the weight of the body) may also be acting If the stresses T_{ij} vary with position application of Newton s second law leads to the equation for the x_1 direction

$$\frac{\partial T_{11}}{\partial x_1} + \frac{\partial T_{12}}{\partial x_2} + \frac{\partial T_{13}}{\partial x_3} + X_1 = \rho f_1$$

where ρ is the density, f_1 is the acceleration, and X_1 the body force component per unit volume along x_1, together with two similar equations for the x_2 and x_3 directions If $f_1 = f_2 = f_3 = 0$ these equations become the equations of equilibrium, and if, further, $X_1 = X_2 = X_3 = 0$, they become the equations of equilibrium in the absence of body forces The preceding equations are important in many branches of elastic theory, and for example provide a starting point in the study of vibrating bodies and of the twisting of cylinders and prisms with cross sections of various shapes See TORSION

The components of strain are specified in a similar way to the stresses There are six independent strain components S_{11}, S_{22}, S_{33}, S_{23}, S_{13}, and S_{12} If as a result of strain, the coordinates of a point x_1 x_2 x_3 become $x_1 + u_1$, $x_2 + u_2$, $x_3 + u_3$ the quantities u_1, u_2, and u_3 are the components of the displacement vector, and the strain components are

$$S_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right)$$

so that, for example,

$$S_{11} = \frac{\partial u_1}{\partial x_1} \qquad S_{12} = \frac{1}{2}\left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right)$$

By eliminating the displacements from these equations the so-called compatibility equations are obtained with three of the type

$$\frac{\partial^2 S_{22}}{\partial x_3^2} + \frac{\partial^2 S_{33}}{\partial x_2^2} = 2\,\frac{\partial^2 S_{23}}{\partial x_2\,\partial x_3}$$

and three of the type

$$\frac{\partial^2 S_{11}}{\partial x_2\,\partial x_3} = \frac{\partial}{\partial x_1}\left(-\frac{\partial S_{23}}{\partial x_1} + \frac{\partial S_{13}}{\partial x_2} + \frac{\partial S_{12}}{\partial x_3}\right)$$

The stresses and strains have so far been denoted by two suffixes This is essential if the methods of tensor analysis are to be applied to elasticity problems but for many purposes a single suffix notation is adequate The change from a two to a one suffix notation for the stresses is simply $T_{11} = T_1$ $T_{22} = T_2$ $T_{33} = T_3$, $T_{23} = T_4$, $T_{13} = T_5$, $T_{12} = T_6$ The change of notation for the strains is $S_{11} = S_1$ $S_{22} = S_2$, $S_{33} = S_3$, $2S_{23} = S_4$ $2S_{13} = S_5$ $2S_{12} = S_6$, the factor 2 is required to make the strains S_4 S_5 and S_6 conform with the usual definition of shear strain (Fig 1b)

Hooke's law generalized. Hooke's law may be generalized to the statement that each stress component is proportional to each strain component equivalent to the six equations

$$T_1 = c_{11}S_1 + c_{12}S_2 + c_{13}S_3 + c_{14}S_4 + c_{15}S_5 + c_{16}S_6$$

$$\cdot \qquad \cdot$$

$$T_6 = c_{61}S_1 + c_{62}S_2 + c_{63}S_3 + c_{64}S_4 + c_{65}S_5 + c_{66}S_6$$

which may be written more concisely as

$$T_q = \sum_r c_{qr}S_r \qquad (3)$$

where the summation extends over $r = 1, 2, 3, 4, 5$ and 6 The elastic constants c_{qr} are termed the stiffnesses, there are altogether 36 of them but they are subject to the reciprocal relations $c_{qr} = c_{rq}$, imposed by thermodynamic requirements, and the number is thus reduced to 21

Additional relations can be derived assuming (i) that the interatomic forces act along the lines joining the centers of atoms in the lattice, (ii) that the atoms are situated at centers of symmetry and (iii) that the lattice is initially at zero stress These relations called Cauchy relations are $c_{23} = c_{44}$, $c_{13} = c_{55}$, $c_{12} = c_{66}$, $c_{14} = c_{56}$, $c_{25} = c_{46}$, $c_{45} = c_{36}$ and if true, would reduce the number of stiffnesses to 15 Experiment shows, however, that they are not true in general, nevertheless, their investigation provides an indication of the extent to which the assumptions (i)–(iii) hold in any particular case

The generalized Hooke's law can also be written to express the strains in terms of the stresses

$$S_q = \sum_r s_{qr}T_r \qquad (r = 1, 2, 3, 4, 5, 6)$$

in which the quantities s_{qr} are the elastic compliances If the six simultaneous Eqs (3) are solved for the strains the compliances are obtained in terms of the stiffnesses as

$$s_{qr} = \Delta c_{qr}/\Delta c$$

where Δc is the determinant

$$\Delta c = \begin{vmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ c_{12} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ c_{13} & c_{23} & c_{33} & c_{34} & c_{35} & c_{36} \\ c_{14} & c_{24} & c_{34} & c_{44} & c_{45} & c_{46} \\ c_{15} & c_{25} & c_{35} & c_{45} & c_{55} & c_{56} \\ c_{16} & c_{26} & c_{36} & c_{46} & c_{56} & c_{66} \end{vmatrix}$$

and Δc_{qr} is the minor determinant obtained by deleting the row and column containing c_{qr} from the determinant Δc

The 21 stiffnesses (or compliances) of the generalized Hooke's law describe the elastic behavior of a material belonging to the triclinic crystal system (see CRYSTALLOGRAPHY) The existence of symmetry elements reduces the number of independent elastic constants in the other crystal systems to the following numbers monoclinic 13, orthorhombic, 9 tetragonal, 7 or 6, trigonal, 7 or 6, hexagonal,

properties are independent of direction the material is isotropic and its elastic behavior is completely described by two independent stiffnesses (or compliances)

The stress-strain relations, referred to the principal axes in the orthorhombic, hexagonal cubic, and isotropic systems, *are given in the accomp*

Stress strain relations

System	Orthorhombic	Hexagonal	Cubic	Isotropic
$T_1 =$	$c_{11}S_1 + c_{12}S_2 + c_{13}S_3$	$c_{11}S_1 + c_{12}S_2 + c_{13}S_3$	$c_{11}S_1 + c_{12}S_2 + c_{12}S_3$	$c_{11}S_1 + c_{12}S_2 + c_{12}S_3$
$T_2 =$	$c_{12}S_1 + c_{22}S_2 + c_{23}S_3$	$c_{12}S_1 + c_{11}S_2 + c_{13}S_3$	$c_{12}S_1 + c_{11}S_2 + c_{12}S_3$	$c_{12}S_1 + c_{11}S_2 + c_{12}S_3$
$T_3 =$	$c_{13}S_1 + c_{23}S_2 + c_{33}S_3$	$c_{13}S_1 + c_{13}S_2 + c_{33}S_3$	$c_{12}S_1 + c_{12}S_2 + c_{11}S_3$	$c_{12}S_1 + c_{12}S_2 + c_{11}S_3$
$T_4 =$	$c_{44}S_4$	$c_{44}S_4$	$c_{44}S_4$	$(c_{11} - c_{12})S_4/2$
$T_5 =$	$c_{55}S_5$	$c_{44}S_5$	$c_{44}S_5$	$(c_{11} - c_{12})S_5/2$
$T_6 =$	$c_{66}S_6$	$(c_{11} - c_{12})S_6/2$	$c_{44}S_6$	$(c_{11} - c_{12})S_6/2$

ing table The equations involving the compliances are completely analogous with S and T interchanged and s_q written for c_q except where $T_q = ½(c_{11} - c_{12})\ S_q$ in which case $S_q = 2(s_{11} - s_{12})\ T_q$

Rochelle salt is an example of an orthorhombic crystal materials which although not crystalline possess the same symmetry and matrix of elastic constants as orthorhombic crystals are said to be orthotropic Wood and plywood are materials of this description and orthotropic elastic theory has also been applied to laminated plastics and reinforced concrete

Single crystal zinc cobalt magnesium and ice are hexagonal materials they are transversely isotropic because the properties are independent of direction in all planes normal to the hexagonal axis

Single crystal copper gold silver nickel and the alkali halides (for example sodium chloride) are important cubic materials The stress strain equations are derived from those of the orthorhombic system by superimposing the condition that the three principal directions are all equivalent This does not mean that the properties are independent of direction for example the compliance s'_{11} in an arbitrary direction is given by

$$s'_{11} = s_{11} - 2(s_{11} - s_{12} - s_{44}/2) \times (a_1{}^2a_2{}^2 + a_2{}^2a_3{}^2 + a_3{}^2a_1{}^2)$$

where a_1 a_2 a_3 are the cosines of the angles between the arbitrary direction and the cubic axes This equation shows that s'_{11} depends on orientation unless $s_{44}/2 = (s_{11} - s_{12})$ See ANELASTICITY PHOTOELASTICITY [R F S H]

Bibliography A E H Love *Treatise on the Mathematical Theory of Elasticity* 4th ed 1927, J F Nye *Physical Properties of Crystals Their Representation by Tensors and Matrices* 1957 F Seitz and D Turnbull (eds) *Solid State Physics*, vol 7 1958 S P Timoshenko *Theory of Elastic Stability* 1936 S P Timoshenko *Theory of Plates and Shells* 1940, S P Timoshenko *Vibration Problems in Engineering* 3d ed 1955 S P Timoshenko and J N Goodier *Theory of Elasticity* 2d ed 1951

Elaterite

A light brown to black naturally occurring carbonaceous substance having specific gravity 0 90–1 05 Elaterite is insoluble in carbon disulfide and infusible It is moderately soft and elastic and on heating leaves only 2–5% fixed carbon Elaterite is thought to be of algal origin and occurs in small quantities in Derbyshire County, England in the Coorong District of South Australia and in Turkestan U S S R See ALBERTITE, IMPSONITE, WURTZILITE, see also ASPHALT AND ASPHALTITE [I A B]

Electret

A solid dielectric possessing persistent dielectric polarization See POLARIZATION (DIELECTRICS) An electret is the analog of a magnet Electrets are made by cooling suitable dielectrics from elevated temperatures in strong electric fields A special class called photoelectrets is produced by the removal of light from an illuminated photoconductor in an electric field

Electrets can be prepared from certain organic waxes and resins (for example carnauba wax) or from ferroelectric crystals or ceramics such as barium titanate Photoelectrets have been prepared from sulfur cadmium and zinc sulfides and anthracene

Electrets are metastable, their polarizations decay slowly after removal of the applied field and more rapidly with increasing temperature Space-charge polarization is the principal mechanism involved in electret formation except for ferroelectric substances See DIELECTRICS, FERROELECTRICS MAGNET [R D W]

Electric distribution systems

That part of an electric power system that supplies electric energy to the individual user or consumer The four general classes of individual users are residential industrial commercial and rural See ELECTRIC POWER SYSTEMS

Distribution systems may be classified as either direct current systems or alternating current systems

Direct current systems Direct current (dc) was the original commercial type of power made available for distribution service and customer use

The dc system in a simple form requires only two wires or conductors As the load increases larger conductors are needed To reduce the need for large conductors Thomas Edison developed the three wire dc system which has 120 volts between each outside wire and the third (neutral) wire and 240 volts between the two outside wires The current in the neutral wire is the sum of two opposing currents for the loads on the two sides of the system The total neutral wire current is therefore never greater than the current in either outside wire and the total copper required is less than that

for two independent two-wire systems Most dc systems have been replaced by ac systems, but a few Edison systems remain in the downtown commercial areas of large cities

Alternating current systems The three-phase alternating current (ac) system (Fig 1) is practically universal although some two phase systems of early vintage continue in operation Single phase branches are used for single phase utilization in residences small stores, and farms Loads are universally connected in parallel to common supply circuits except for some street lights which operate in series on a constant current system The series circuit requires its own regulating transformer to supply a constant current through the lamps by raising the voltage when the lamps are added to the circuit

Primary voltages. Power leaving the distribution substations is generally at 2400 volts line to neutral and 4160 volts line to line (conventionally [illegible] volts [illegible] and [illegible] wire systems, single phase loads are connected line-to neutral These primary voltages are stepped down through pole-mounted distribution transformers to the lower secondary voltage to supply individual consumers

Large light and power loads are usually supplied by three phase primary voltage which is stepped down through a transformer to the user s own distribution system The larger industrial loads are supplied from voltages ranging from 13 200 to 132 000 volts These are three phase three wire systems, because all loads are three-phase

Secondary voltages Secondary voltages usually correspond to utilization voltages Residences and most farms are supplied by 120/240 volt single phase three wire systems Commercial and small light and power loads are supplied by 120/208 volt or 277/480 volt three phase, four wire systems

The secondary voltage is used sometimes on lines supplying multiple street lamps in addition to service to customers Street lighting systems using multiple street lamps are increasing in acceptance relative to series lamp systems A multiple street lamp system has all lamps connected in multiple (or parallel) and energized by a constant potential service usually 120 volts

From the viewpoint of system design, service to the consumer is thought of in two terms good voltage and good continuity

Good voltage. Utilization voltages vary from company to company but a voltage variation of less than 5% at the customer's meter is common for residential and commercial loads To achieve this result voltage is regulated at a substation bus or at an individual feeder at the substation, or supplementary regulation is provided at one or more points along the line Voltage regulators are discussed later in this article

Good continuity The term good continuity is not usually expressed quantitatively because of the difficulty of doing so A brief explanation is given here Large industrial and commercial loads are supplied with some form of duplicate feeds or power supply Downtown business areas of large cities are supplied from 3 phase 120/208 volt networks and 277/480-volt networks are being introduced These networks are fed from a plurality of primary feeders stepping down through transformers then through automatic reclosing circuit breakers to the secondary grid formed by the cables following the streets and avenues and interconnected at intersections Residential and farm areas do not have duplicate feeds to consumers Reliance is placed upon sectionalizing the system by fuses circuit breakers and manual disconnecting devices to reduce the extent of an outage and the time taken to locate and repair the difficulty and restore service

Primary feeders Power is carried by primary feeders from distribution substations to the load areas where the consumers are located In some areas of some cities the 2400/4160 volt primary feeders are interconnected to form a grid and are operated somewhat analogously to the secondary 120/208 volt networks

Secondary mains This element of electric distribution is a low voltage system which connects the secondary windings of distribution transformers to the customers' services The transformers are supplied by primary feeders and are mounted on poles in building vaults and in some instances in curb vaults

Two general schemes of connections are used radial and network In radial schemes the mains are energized either from one end or near the center of a conductor run to supply general lighting and small power These mains are usually single-phase three-wire operating at 120 volts line to-

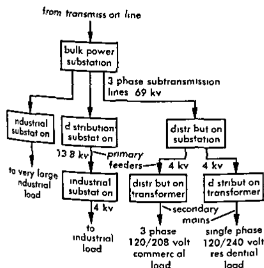


Fig 1 Typical three phase distribution system extends from bulk power source to consumers load switches Typical voltages are indicated Single line diagram is used for simplicity

tral and 240 volts line to line (120/240 volts) The secondaries are tapped for service to customers' premises

The network scheme of connections comprises the secondary mains which serve as a tie between transformers on different feeders in some cases forming an apparent loop In the event of failure of a primary feeder or a transformer the other feeders and transformers take a share of the load This scheme is used in cities where the electric load per city block is high Service is commonly 120 208 volt three phase four wire providing 120 volts line-to neutral for lighting and 208 volts for three phase power service Connections are made on the network conductors to run service to the customers

Overhead construction Overhead installation is general in residential industrial and farm areas This construction utilizes southern yellow pine poles treated with pentachlorophenol in the eastern half of the nation and western red cedar in the western half Conductors generally have weather proof coverings in urban areas but an increasing amount of bare conductors is now being used Since World War II much aluminum has been used in place of copper

Underground construction Commercial areas and certain main thoroughfares and streets traversed by several circuits usually have underground construction Some residential real estate developers are now paying utilities an extra charge for buried cable installations Underground construction is usually a system of concrete or fiber ducts and manholes *Impregnated* wood pulp paper and natural and synthetic rubber compounds are generally used as conductor insulation Paper cables are protected by a *lead sheath* Rubber compounds also used lead sheaths but many are now being covered with a polymerized chlorobutadiene rub-

place to place The worst circumstances occur where stray currents from dc railway or Edison three wire dc systems enter and leave cable sheaths or equipment Stray 60-cycle currents do not produce significant corrosion as a rule

Bare copper wire, galvanized supports or equipment, cinder filled ground, electrochemical cells caused by differing environments anaerobic bacteria in poorly aerated soils, organic acids from wood ducts, sewage, rotting vegetation acetic acid, carbon dioxide, and alkalies frequently contribute significantly to corrosion

Corrosion mitigation. When corrosion is caused by dc stray currents, mitigation begins at the source of the stray current Tracks for electric traction systems should be bonded rail to rail to reduce voltage drops Leakage should be eliminated from line to ground on dc Edison three-wire systems Current drainage bonds from cable sheath to railway negative bus are also helpful

Galvanic currents. Electrochemical attacks are mitigated by corrosion preventive coverings and grease over sheaths or pipes by substitution of clean sand for cinders and by cleaning the ducts and manholes Cathodic protection may also be used This is the application of currents from a separate source of a magnitude adequate to prevent discharges from anodic spots (see Fig 2) Currents can also be generated with galvanic anodes like aluminum zinc, and magnesium and give cathodic protection in some situations Bare copper for neutrals grounds and bonding connections are protected by tinning

Distribution equipment. Apparatus for transforming fault clearing sectionalizing switching circuit reclosing and regulating operations will be outlined

Transformers Pole mounted single phase distribution service transformers are standardized in 5 to 167 kva sizes It is general practice to install a primary fuse and lightning arrester with each transformer The fuse is often mounted inside the trans-

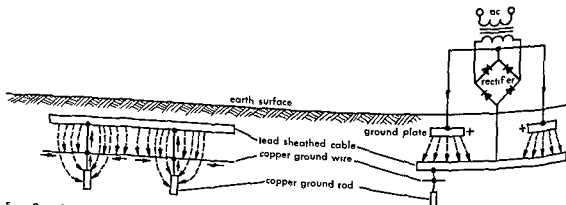


Fig 2 Corrosion and *cathodic* protection Lead sheath is grounded by copper wire and rod but corrodes because of dissimilarity of metals in the soil Arrows (dashed) indicate flow of current from lead (anodic) through soil to copper (cathodic) and by short arrows on copper back to lead completing a galvanic circuit At right cathodic protection is provided by a rectifier which reverses the current (dashed arrows) from the ground plate to lead sheath

former tank and a current thermal trip circuit breaker is added on the secondary side The fault self-clearing characteristics of overhead conductors are relied upon where secondary 120/240-volt breakers are not used A primary fuse employs an expulsion action generated by fault currents to clear a faulted transformer *See* FUSE, ELECTRIC

The same primary fuses called primary fuse cutouts are also used to subdivide or sectionalize the primary circuit into parts An automatic reclosing pole-mounted circuit breaker is also used for circuit sectionalizing The two are often combined in the same circuit The source or substation end of the primary feeder is protected by a station type automatic circuit breaker This breaker is usually the air break type with magnetic blow out chutes Formerly oil circuit breakers were in universal use *See* CIRCUIT BREAKER

Voltage regulating apparatus Equipments for regulating voltage are generally located at the source or substation end of the feeder They may be located to regulate a bus supplying several feeders or they may be located in the individual feeders Bus regulators are now usually a part of the power transformer and are of the tap changing type although separate installations are common and sometimes of the induction type *See* VOLTAGE REGULATOR

Feeder voltage regulators have generally been of the induction type but the tap changing type is now being used extensively Pole mounted tap changing regulators are widely used as supplementary regulators in the vicinity of the load

Capacitors Formerly known as static condensers capacitors are also used to improve voltage by the neutralization of the voltage drop caused by the lagging or low power factor load currents They are used switched and unswitched Switched capacitors are used at substations to influence voltage levels at the substation bus Unswitched capacitors are connected continuously to a circuit

System engineering Of the two aspects of engineering technical and economic the economic is dominant in distribution system engineering because of the maturity of the technical art

The relatively low voltages used in distribution systems has necessitated a large number of circuits to carry the load This has caused circuit congestion on streets and the location of new substations has been a problem because of space restrictions The solution is to use a higher voltage on the distribution feeders and sometimes the subtransmission lines also This results in fewer circuits and substations and also results in more load carried per circuit with consequent larger substations

Higher voltage circuits produce problems of tree interference and greater outages seen by a customer when compared with a lower voltage circuit Fortunately the shorter circuit lengths at greater load densities reduce the outage rates and some further reduction can be made by use of sectionalizing The important consideration is to foresee the ultimate load growth requiring the higher voltage and to introduce it as early as possible and thereby reduce the costs associated with conversion of existing circuits and equipment to the higher voltage Modern real estate developments of hundreds and thousands of homes in a given area within a period of one to three years have facilitated the use of higher distribution voltages at the beginning *See* ELECTRIC POWER SUBSTATION

[D K B]

Bibliography American Standards Association [illegible]

Handbook for Electrical Engineers 9th ed 1957

Electric energy measurement

The measurement of the integral with respect to time of the power in an electric circuit

The absolute unit of measurement of electric energy is the joule The joule however is too small (1 watt second) for use in commercial practice and the more common unit is the watt hour (3.6×10^3 joules)

The most common measurement application is in the utility field where it is estimated that over 560 000 000 000 kilowatt hours were measured and sold during 1957 to industry and residential consumers in the United States

Electric energy is one of the most accurately measured commodities sold to the general public Many methods of measurement with different degrees of accuracy are possible The choice is dependent upon the requirements and complexities of the problems Basically measurements of electric energy may be classified into two categories the first dealing with direct current power and the second dealing with alternating current power The fundamental concepts of measurement are however the same for both

Methods of measurement There are two types of methods of measuring electric energy (1) electric instruments and timing means and (2) electricity meters

Electric instruments and timing means These methods make use of conventional procedures for measuring electric power and time (*see* ELECTRIC POWER MEASUREMENT TIME INTERVAL MEASUREMENT) The required accuracy of measurement dictates the type and quality of the measuring devices used (for example portable instruments laboratory instruments potentiometers stop watches chronographs electronic timers) Typical methods are listed below

1 Measurement of energy on a direct current circuit by reading the line voltage and load current at regular intervals over a measured period of time The frequency of reading selected (such as

AGE MEASUREMENT

In electric energy measurements the losses in the instruments must be considered Unless negligible from a practical standpoint they should be deducted from the total energy measured If the voltmeter losses are included in the total energy measured then

$$\text{watt hours} = \frac{(EI - E^2/R)t}{3600}$$

where E is the average line voltage (volts) I is the average line current (amperes) R is the voltmeter resistance (ohms) and t is the time (seconds)

2 Measurement of energy on a direct current circuit by controlling the voltage and current at constant predetermined values for a predetermined time interval This method is common for controlling the energy used for a scientific experiment or for determining the accuracy of a watt hour meter For best accuracy potentiometers and electronic timers are desirable

3 Measurement of energy on an alternating current circuit by reading the watts input to the load at regular intervals over a measured period of time This method is similar to the first except that the power input is measured by a wattmeter

4 Measurement of energy on an alternating current circuit by controlling the voltage current and watts input to the load at constant predetermined values This method is similar to the second except that the power input is measured by a wattmeter A common application of this method is to determine the standard of measurement of electric energy the watt hour

5 Measurement of energy by recording the watts input to the load on a linear chart progressed uniformly with time This method makes use of a conventional power record produced by a recording wattmeter The area under the load record over a period of time is the energy measurement

Electricity meters These are the most common devices for measuring the vast quantities of electric energy used by industry and the general public The same t
for e
elect
mear
of electric energy measurement

A single meter is sometimes used to measure the energy consumed in two or more circuits Multi stator meters are generally required for this purpose Totalization is also accomplished with fair accuracy if the power is from the same voltage source by paralleling secondaries of instrument current transformers of the same ratio at the meter Errors can result through unbalanced loading or use of transformers with dissimilar characteristics

Watt hour meters are generally connected to measure the losses of their respective current circuits These losses are extremely small compared to the total energy being measured and are present only under load conditions

Other errors result from the ratio and phase angle errors in instrument transformers With modern transformers these errors can generally be neglected for commercial metering If considered of sufficient importance they can usually be compensated for in adjusting the calibration of the watt hour meter

For particularly accurate measurements of energy over short periods of time portable standard watt hour meters may be used *See* WATT HOUR METER

Quantities other than watt hours Included in the field of electric energy measurement are the measurements of demand var hours and volt ampere hours

Demand The American Standards Association (ASA) defines the demand for an installation or system as "the load which is drawn from the source of supply at the receiving terminals averaged over a suitable and specified interval of time Demand is expressed in kilowatts kilovolt amperes amperes kilovars and other suitable units" (ASA C12 1941)

This measurement provides the user with information as to the loading pattern or the maximum loading of equipments rather than the average loading recorded by the watt hour meter It is used by the utilities as a rate structure tool

Var hour ASA defines the var hour (reactive volt ampere hour) as the "unit for expressing the integral of reactive power in vars over an interval of time expressed in hours" (ASA C12-1941)

This measurement is generally accomplished by making use of reactors or phase-shifting transformers to supply to conventional meters a voltage equal to but in quadrature with the line voltage

Volt ampere hour This is the unit for expressing the integral of apparent power in volt amperes over an interval of time expressed in hours

Measurement of these units is more complicated than for active or reactive energy and requires greater compromises in power factor range accuracy or both Typical methods include

1 Conventional watt hour meters with reactors or phase shifting transformers tapped to provide an in phase line voltage and current relationship applied to the meter at the mean of the expected range of power factor variation

2 A combination of a watt hour and a var hour meter mechanically acting on a rotatable sphere to add vectorially watt hours and var hours to obtain volt ampere hours volt ampere demand or both

Measurement of volt ampere hours is sometimes preferred over var hours because it is a more direct measurement and possibly gives a more accurate picture of the average system power factor This would not necessarily be true however where simultaneous active and reactive demand are measured and recorded *See* ELECTRICAL MEASUREMENTS [C.R.S.]

Bibliography American Standards Association *Code for Electricity Meters* ASA C12 1941 Edison Electric Institute *Electric Metermen's Handbook* 6th ed 1950 L E Janetos and J J Hall *Precise Determination of the Watthour* AIEE CP56-899

Electric field

A condition in space in the vicinity of an electrically charged body such that the forces due to the charge are detectable. An electric field (or electrostatic field) exists in a region if an electric charge at rest in the region experiences a force of electrical origin. Since an electric charge experiences a force if it is in the vicinity of a charged body, there is an electric field surrounding any charged body.

Field strength. The electric field intensity (or field strength) **E** at a point in an electric field has a magnitude given by the quotient obtained when the force acting on a test charge q placed at that point is divided by the magnitude of the test charge q. Thus it is force per unit charge. A test charge q is one whose magnitude is small enough so it does not alter the field in which it is placed. The direction of **E** at the point is the direction of the force **F** on a positive test charge placed at the point. Thus **E** is a vector point function, since it has a definite magnitude and direction at every point in the field, and its defining equation is

$$\mathbf{E} = \frac{\mathbf{F}}{q} \qquad (1)$$

Principle of superposition. As applied to electric fields this principle states that the total **E** at a point P due to the combined influence of a distribution of point charges is the vector sum of the electric field intensities that the individual point charges would produce at P if each acted alone. Thus, using the rationalized mks system of units,

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{n} \frac{q_i}{r_i^2} \qquad (2)$$

where $\epsilon_0 = 8.85 \times 10^{-12}$ coul²/newton m² is the permittivity of empty space, q_i is the ith charge (in coulombs) in the distribution, and r_i is the distance in meters from q_i to P. The units of **E** in the mks system are newtons/coulomb, which are the same as volts/meter. *See* GAUSS' THEOREM. A common method of solving for **E** in a particular known distribution of charges is to evaluate the vector sum in Eq. (2). In many cases, however, Gauss' theorem affords a more powerful and convenient method.

Electric displacement. Electric flux density or electric displacement **D** in a dielectric (insulating) material is related to **E** by either of the equivalent equations

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} \quad \text{and} \quad \mathbf{D} = \epsilon\mathbf{E} \qquad (3)$$

where **P** is the polarization of the medium and ϵ is the permittivity of the dielectric, which is related to ϵ_0 by the equation $\epsilon = k\epsilon_0$, k being the relative dielectric constant of the dielectric. In empty space $\mathbf{D} = \epsilon_0\mathbf{E}$. The units of **D** are coulombs/meter².

In addition to electrostatic fields produced by separations of electric charges, an electric field is also produced by a changing magnetic field. The relationship between the **E** produced and the rate of change of magnetic flux density $d\mathbf{B}/dt$ which produces it is given by Faraday's law of induced emfs in the form

$$\oint \mathbf{E} \cdot d\mathbf{s} = -\int_A \frac{d\mathbf{B}}{dt} \cdot d\mathbf{A} \qquad (4)$$

where $d\mathbf{s}$ is a vector element of path length directed along the path of integration in the general sense of **E**. Thus $\oint \mathbf{E} \cdot d\mathbf{s}$ is the emf induced in this closed path of integration. The area of the surface bounded by the path of integration is A, and the direction of $d\mathbf{A}$, an infinitesimal vector element of this area, is the direction of the thumb of the right hand when the fingers encircle the path of integration in the general sense of **E**. The right side of Eq. (4) is seen to be the negative of the time rate of change of the magnetic flux linking the path of integration chosen for the left side.

In an electrostatic field $\oint \mathbf{E} \cdot d\mathbf{s}$ is always zero. *See* CHARGE, ELECTRIC; INDUCTION, ELECTROMAGNETIC; POTENTIAL, ELECTRIC. [R.P.WI.]

Bibliography: R. P. Winch, *Electricity and Magnetism*, 1955.

Electric furnace

An enclosed space heated by electric power. The furnace may be in such forms as a refractory crucible, a large tiltable refractory basin with a capacity of 100 tons and a removable roof, or a long insulated chamber equipped with a continuous conveyor. The heat is provided by an arc to the charge or melt (direct arc furnace) or by an arc between electrodes (indirect arc furnace), an electric current conducted to the melt by a pair of electrodes (resistance or submerged arc furnace), an electric current induced into the metal from a surrounding coil, or by radiation from an electrical resistance near the charge. Because the source of heat is nonchemical, electric furnaces are especially desirable in melting alloys of controlled composition. Temperature is also readily controlled. The arc furnace may be used to smelt ores or to refine metals or alloys. Induction furnaces are widely used to melt alloys for casting. Furnaces with hearth resistors are used for operations below melting temperatures, such as annealing. *See* ARC HEATING; HEAT TREATMENT (METALS AND ALLOYS); HEATING, ELECTRIC; INDUCTION HEATING; *see also* FURNACE CONSTRUCTION; KILN. [F.H.R.]

Electric organ (biology)

An organ consisting of rows of plates (electroplaques) which produce an electrical discharge. Electric organs are known in only seven families of marine and fresh-water electric fish. They arise bilaterally from modified muscle fibers, electroplaques. Special anatomical and physiological features of the latter permit series summation of the electromotive forces (emfs) generated by the individual cells. Unlike other electrogenic tissues, such as muscles or nerves, electric organs can develop appreciable voltages in the surrounding fluid. The columnar arrays of several hundred to several thousand electroplaques in series in the strongly

electric fish *Electrophorus Malapterurus* and *Torpedo* are paralleled so that the electric organs of these fish can generate considerable current at high voltage *T nobiliana* may develop up to 50 amp on short circuit and *Electrophorus* about 1 amp The electric organs form a major part of the body of the strongly electric fish (Fig 1) The discharges controlled by the nervous system are used for offense and defense The weakly electric fish have only a few columns of series arrays and relatively few electroplaques in each column However many of the species emit pulses of low voltage more or less continuously and regularly The form frequency and regularity of the discharges often are characteristic for a given species These pulses may serve as power components in an electrical guidance system the receptors for which must detect extremely small changes in the conductance of the water The numerous anatomical and functional varieties of electric organs provide much general information regarding the nature of bioelectric activities *See* BIOELECTRIC MODEL BIOPOTENTIALS AND ELECTROPHYSIOLOGY

Elasmobranch marine electric fish comprise the four genera of Torpedinidae (electric torpedoes or rays) and the four of the electric skates (Rajadae) Both families are widely distributed in the oceans The only known teleost marine electric fish is the stargazer (*Astroscopus*) found off the Atlantic Coast from the southern United States to Brazil All fresh water electric fish are tropical The American family Gymnotidae includes the electric eel *Electrophorus* as well as the numerous species of weakly electric knifefish *Malapterurus Gymnarchus*, and another family, the Mormyridae which includes many species are found in Africa

Electroplaque. Different groups of muscles have been transformed into electric organs in different forms (Table 1) The gross anatomy of the organs and of their nerve supply therefore shows considerable variety The structure of the individual electroplaques however, exhibits a common pattern in most electric fish The cells are relatively thin They appear in most species as two large wafer like roughly circular or rectangular surfaces. However in some species of knifefishes the electroplaques are elongated cylinders

Electroplaques are usually innervated by a number of axons which may derive from different nerve trunks Only one of the major faces is innervated. In most forms the axon terminals are applied directly to this surface and branching profusely they make numerous synaptic connections over the entire innervated surface In some fish (*Malapterurus* mormyrids) innervation is at the tips of stalk processes which are produced from one major face Cylindrical electroplaques are innervated at one end The innervation (see table) defines the uniform orientation of all the electroplaques in a columnar series array In some species all the series columns are identically oriented but in others there is a patterned difference Thus of the four columns which comprise the electric organ in *Gymnotus carapo* one has its electroplaques innervated on their rostral surface the other three on their

Table 1 Anatomy of electroplaque in several electric fish

Species	Origin (muscle)	Innervation*	Orientation	Dimensions R C	Dimensions D V	Dimensions M L†	No in columns	No of columns per side
Torpedo nobiliana	Branchial	V	D V	8 mm	10 μ	8 mm	1000	1000
Narcine brasiliensis								
Main organ	Branchial	V	D V	4 mm	10 μ	4 mm	500	400
Accessory organ	Branchial	D	Oblique	4 mm	20 μ	4 mm	200	10
Raia clavata	Skeletal	R	R C				200	12
Astroscopus y-graecum	Ocular	D	D V	10 mm	50 μ	10 mm	200	20
Electrophorus electricus	Skeletal	C	R C	200 μ	1 mm	15 mm	6000	70
Eigenmannia virescens	Skeletal	C	R C	2 mm	200 μ	200 μ		5
Sternopygus elegans	Skeletal	C	R C	1 mm	60 μ	60 μ		15
Gymnotus carapo	Skeletal	R and C	R C	200 μ	500 μ	500 μ	80	4
Sternarchus albifrons	?	?	R C					
Gnathonemus compressirostris	Skeletal	C	R C	50 μ	10 mm	5 mm	100	2
Mormyrus rume	Skeletal	C	R C	50 μ	10 mm	5 mm	100	2
Gymnarchus niloticus	Skeletal	C	R C	100 μ	100 μ	100 μ	140	4
Malapterurus electricus	?	C	R C	40 μ	1 mm	1 mm	3000	1500

* Abbreviations are R rostral C caudal D dorsal and V ventral † Medial lateral

caudal surface In *Narcine brasiliensis* and probably in other close relatives but not in *T nobiliana* and other species of *Torpedo*, there is an accessory organ whose electroplaques are dorsally innervated although the main organ conforms to the general torpedine pattern of innervation on the ventral surface

Neural and chemical stimulation Several kinds of functional differences characterize the properties of the electroplaques of different forms Only the innervated surface is reactive in the electroplaques of marine electric fish The electrogenic activity is evoked by stimulating the nerve or by chemical agents but not by direct electrical stimuli that is the cell is electrically inexcitable The responses have the properties of depolarizing postsynaptic potentials decreasing the steady state internal negativity (the resting potential) Thus during a discharge the exterior of the innervated surface becomes negative (Fig 2) Series summation of voltages therefore takes place by disappearance of one of two opposing and equal batteries The maximum contribution of each electroplaque is approximately the emf of the resting potential about 50–80 mv per electroplaque in different species This contribution is maximized by the relatively low resistance of the inactive noninnervated surface which reduces the *IR* drop of the action current across this face The responses last only a few milliseconds in torpedines and *Astroscopus* but in Rajadae the duration is much longer (Table 2) The synaptic membrane is cholinoceptive and its pharmacological properties in general resemble those of muscle end plates

External stimulation In addition to the neurally evoked postsynaptic potential all fresh water fish thus far studied electrophysiologically have a second electrogenic activity This component is electrically excitable and may be triggered by an electrical stimulus as well as by the postsynaptic potential of normal activity The response (spike) is *a temporary reversal of the membrane potential*

Table 2. Electroplaque discharge and response characteristics in several electric fish

Species	Discharge			Response			
						Duration msec	
	Amplitude volts	Form	Frequency	Amplitude mv	Type*	Post-synaptic potential	Spike
Torpedo nobiliana	60	Monophasic	Repetitive on excitation	Max 80	1	5	None
Narcine brasiliensis							
Main organ	30	Monophasic	Repetitive on excitation	Max 80	1	5	None
Accessory organ	0 5	Monophasic	Repetitive on excitation	Max 80	1	5	None
Raja clavata	4	Monophasic	Repetitive on excitation	Max 80	1	25	None
Astroscopus y-graecum	7	Monophasic	Repetitive on excitation	Max 80	1	5	None
Electrophorus electricus	700	Monophasic	Repetitive on excitation	Min 100	2	2+	2+
Eigenmannia virescens	1	Monophasic positive direct current	250/sec	Min 100	2	1	2
Sternopygus elegans	1	Monophasic positive direct current	50/sec	Min 100	2	2	10
Gymnotus carapo	1	Triphasic	50/sec	Min 100	3	1 5	1
Sternarchus albifrons	1	Diphasic	750/sec	Min 100	3		
Gnathonemus compressirostris	10	Diphasic	Variable	Min 100	4		0 2
Mormyrus rume	12	Diphasic	Variable	Min 100	4		5
Gymnarchus niloticus	Low	Monophasic	300/sec	?	?		
Malapterurus electricus	300	Monophasic	Repetitive on excitation	Min 100	4		2

*Response types 1 electrically inexcitable electroplaques which produce only a postsynaptic potential and only o … nervat … inners … from t … in pr … urlares produce spikes

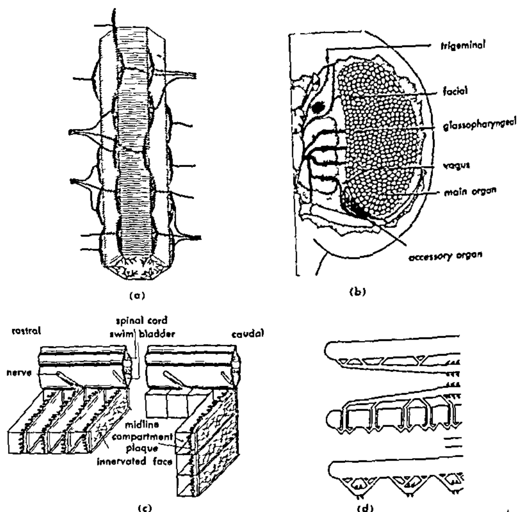


Fig 1 Samples of organ and electroplaque structure (a) Column of electroplaques in series array, representing essentially the arrangement in the torpedine electric fishes and in *Astroscopus* (b) Dorsal view of innervation which applies to *Torpedo* and main organ of *Narcine* innervation is by individual nerve fibers to ventral surface of each electroplaque entering four different points of the periphery and supplying a limited area of the surface In *Astroscopus* and the accessory organ of *Narcine* innervation is on the rostral surface and nerve supply is more complicated Figure also applies to *Torpedo* except that accessory organ is absent (c) Diagrammatic view of series and parallel arrays of electroplaques in the electric eel A somewhat similar series parallel arrangement occurs in other electric fish in which one surface is innervated In *Raia* innervation is on rostral surfaces (d) The mormyrid electroplaques are innervated on one or several stalk processes which form from branches that arise in the caudal surface of each electroplaque In some branches penetrate through the electroplaque body and innervation is then ahead of the electroplaque In *Malapterurus* there is only a single stalk which arises from the center of the caudal face of the electroplaque

to internal positivity This response like the postsynaptic potential is confined to the innervated surface in *Electrophorus* and some other gymnotids and probably also in *Gymnarchus* The contribution of each electroplaque in a series array therefore is a monophasic potential considerably larger than the resting potential representing a temporary inversion of the polarity of one of the two batteries the active membrane In other knife fishes, in *Malapterurus*, and in mormyrids the uninnervated surface too, is electrically excitable and produces a spike In *G carapo* the response of this surface is triggered by the spike of the innervated face which is itself triggered by the neurally evoked depolarizing postsynaptic potential The individual electroplaques therefore contribute a diphasic potential smaller than the full emf of the active cell The exterior of the innervated surface, however, obeys Pacini's rule because activity and external negativity always start at the innervated face The postsynaptic potentials of the gymnotid electroplaques are all brief and are generated by cholinoceptive membrane See ELECTROPHYSIOLOGY (HEART)

In electroplaques which are innervated at the tip of one (*Malapterurus*) or of several processes (mormyrids) synaptic excitation is distant from the major surfaces and has not yet been adequately studied It triggers an electrically excitable response which propagates along the single or branched stalk and invades the main cell surface Both of the major faces of the electroplaques are electrogenic in *Malapterurus* and the mormyrids The spike of the anterior (uninnervated) surface of *Malapterurus* electroplaques arises earlier is slightly larger, and lasts much longer than the response of the posterior face from which the stalk emerges The net contribution of each electroplaque is therefore only slightly diminished by the temporarily opposing emf of the posterior face However the potential of the organ discharge thus breaks Pacini s rule, because the posterior surface from which the stalk that receives the innervation emerges is then always *positive to the rostral face* The spikes of the two faces in many mormyrids are about equal except that one starts earlier The contribution to the total discharge is a brief diphasic potential The spikes may be so separated in time that the peak to peak amplitude of the external potential is almost the sum of the spikes of each face Except for small prefatory potentials in some species *the externally recorded discharge begins as* negativity of the innervated surfaces of the electroplaques Thus the discharges obey Pacini's rule in their main part In some species, however the spike of the uninnervated surface is longer lasting (Fig 2) The net discharge, like that in *Malapterurus*, therefore constitutes negativity of the uninnervated surface, contrary to Pacini's rule.

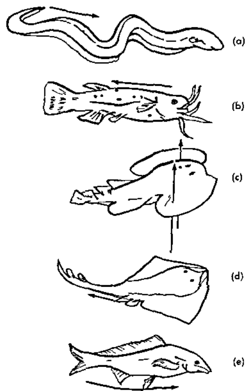


Fig 2 External appearances of various electric fish and the direction of current flow during discharge (a) *Electrophorus* (b) *Malapterurus* (c) *Torpedo* (d) *Raia* (e) Mormyrid Head of arrow indicates the positive pole on external recording, or the direction of current flow inside the fish In all except *Malapterurus* the negative pole corresponds to the side on which the electroplaques are innervated This is Pacini s rule which is also broken in the case of some mormyrids, in which the innervation is caudal but the discharge is head negative as in *Malapterurus*

Ultrastructure Electron microscopy reveals that the unreactive uninnervated surfaces of marine electroplaques or of *Electrophorus* cells as well as both electrically excitable surfaces of *Malapterurus* electroplaques possess canalicular networks Those surfaces immediately in contact with nerve fibers can be demarcated histochemically by staining for cholinesterase However neither the remarkable variety of functional differences among different electroplaques nor even the differences among different parts of one cell are revealed by microscopic methods This fact indicates that the functional differences arise from variations of molecular structures in the cell membrane This can not as yet be resolved by electron microscopy

Innervation The nerve fibers innervating electroplaques are axons of motoneurons whose discharges in turn are controlled from higher centers In some fish a distinct command nucleus has been located in the medulla Only one pair of giant motoneurons lying bilaterally in the medulla innervates the electroplaques of *Malapterurus* The single axon of each side therefore branches enormously to supply the several million cells of the organ

Nevertheless the organ discharge in all electric fish is a highly synchronized activity of all the electroplaques despite propagational delays that must be involved in distributing excitation by the impulses in the various portions of the efferent pathways In *Malapterurus* these delay paths involve *only the conduction times in the different branches* of each of the two axons but in other fish the delays include conduction times in the many different nerve fibers and in the intraspinal conduction paths from the command nucleus to the motoneurons Compensatory delay mechanisms of longer duration for those parts of the system with the shorter conduction delays equalize the time differences and synchronize the discharges The compensatory delay can only be peripheral in *Malapterurus*, but is central and peripheral in gymnotids

Control mechanism and sensory system The continuously discharging fishes exhibit two other

in avoidance of obstacles and perhaps also for species identification In some species of fish the fre

quencies of the emitted pulses are higher than 1000 per second extremely regular and inde pendent of peripheral stimuli The form fre quency and regularity of the pulses are characteris tic for each of the gymnotid species thus far studied but among the mormyrids the responses appear to be restricted to only a few varieties

A change of the discharge frequency can be pro duced in the gymnotids by cooling the head alone which is the site of the command nucleus This nu cleus appears to receive influences from the pe riphery in some species only In them peripheral stimuli cause a transient change in frequency of the discharges In other species however the fre quency is not changed even by strong stimuli In the mormyrids the frequency is highly variable de pending upon the activity or excitation of the fish Obviously the organs which produce continuous high frequency discharges require the expenditure of considerable energy to overcome the ionic mix ing which occurs during activity However this problem has not yet been investigated in detail

All species of the continuously emitting fish that have been studied are sensitive to changes in the conductance of the water Presumably the fish sense the altered electric field of their discharges but the sensory receptor cells have not yet been identified It is inferred that they are distributed along the body as part of the lateral line system In *G carapo* the afferent axons form part of the facial nerve of that system and terminate in the cranial portion of the neuraxis from which the sensory input can initiate motor activity *See* SENSE ORGAN [H GT]

Electric power generation

The production of bulk electric power for indus trial residential and rural use Although limited amounts of electricity can be generated by many means including chemical reaction (as in bat teries) and engine driven generators (as in auto mobiles and airplanes) electric power generation generally implies large scale production of electric power in stationary plants designed for that pur pose The generating units in these plants convert energy from falling water coal natural gas oil and nuclear fuels to electric energy Most electric generators are driven either by hydraulic turbines for conversion of falling water energy or by steam turbines for conversion of fuel energy Electric power generating plants are normally intercon nected by a transmission and distribution system to serve the electric loads in a given area or region *See* GENERATOR ELECTRIC PRIME MOVER

An electric load is the power requirement of any device or equipment that converts electric energy into light heat or mechanical energy or otherwise consumes electric energy as in aluminum reduc tion or the power requirements of electronic and control devices The total load on any power system is seldom constant, rather it varies widely with hourly weekly monthly or annual changes in the requirements of the area served The minimum system load for a given period is termed the base load or the unity load factor component Maxi mum loads resulting usually from temporary con ditions, are called peak loads Electric energy can not be stored in large quantities, therefore the operation of the generating plants must be closely coordinated with fluctuations in the load

Actual variations in the load with time are recorded and from these data load graphs are made to forecast the probable variations of load in the future A study of hourly load graphs (Fig. 1) indicates the generation that may be required at a given hour of the day, week or month A study of annual load graphs indicates the rate at which new generating stations must be built Load graphs are an inseparable part of utility operation and are the basis for decisions that profoundly affect the finan cial requirements and over all development of a utility

Generating plants Often termed generating sta tions these plants contain apparatus that converts some form of energy to electric energy in bulk Three significant types of generating plants are hydroelectric, fossil fuel electric and atomic elec tric

Hydroelectric plant This type of generating plant utilizes the potential energy released by the weight of water falling through a vertical distance called head Ignoring losses the power obtainable from falling water is

$$\begin{pmatrix}\text{Horse}\\ \text{power}\end{pmatrix} = \frac{\begin{pmatrix}\text{Quantity of water in}\\ \text{cubic feet per second}\end{pmatrix}\begin{pmatrix}\text{Vertical head}\\ \text{in feet}\end{pmatrix}}{8\ 8}$$

$$\begin{pmatrix}\text{Kilo}\\ \text{watts}\end{pmatrix} = 0\ 746 \times \text{power in horsepower}$$

A plant consists basically of a dam to store the water in a forebay and create part or all of the head a penstock to deliver the falling water to the turbine a hydraulic turbine to convert the hy draulic energy released to mechanical energy an alternating current generator (alternator) to con vert the mechanical energy to electric energy and all accessory equipment necessary to control the power flow, voltage and frequency and to afford the protection required (Fig 2)

Fossil fuel electric plant This type utilizes the energy of combustion from coal oil or natural gas A typical large plant consists of fuel process ing and handling facilities a combustion furnace and boiler to produce and store the steam a steam turbine an alternator and the accessory equipment required for plant protection and for control of voltage frequency and power flow A steam plant can frequently be built near a convenient load center provided an adequate supply of cooling water and fuel is available and is readily adapt able to either base loading or peak loading

Atomic electric plant In this type of plant one or more of the nuclear fuels are utilized in a suitable type of nuclear reactor which takes the place of the combustion furnace in the typical steam elec

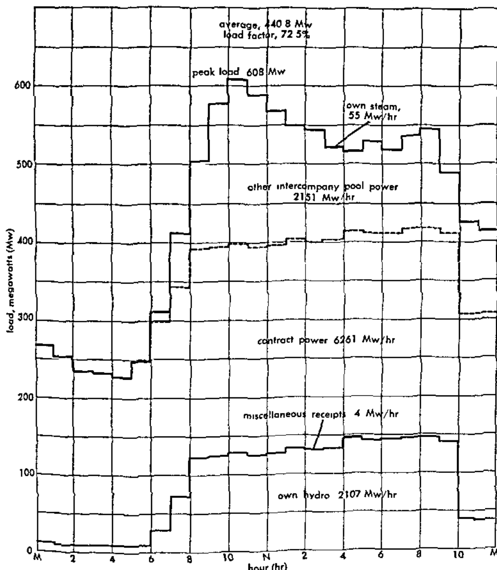


Fig 1 Load graph indicates net system load of a metropolitan utility for typical 24 hour period (mid night to m dnight), totaling 10 578 Mw hr

tric plant The heat exchangers and boilers (if not combined in the reactor), the turbines and alternating current generators, complete with controls accessories and auxiliaries, make up the atomic electric plant These plants are in the research and development stage with a few large scale fission reaction plants in operation or under construction They promise to be a major factor in the energy supply of the future Fusion reaction plants are in the early research and development stage Direct conversion from nuclear reaction energy to electric energy on a commercial scale is a future possibility but is not economically feasible at present

Power plant circuits Both main and accessory circuits in power plants can be classified as follows

1 Main power circuits to carry the power from the generators to the step up transformers and on to the station high voltage terminals

2 Auxiliary power circuits to provide power to the motors used to drive the necessary auxiliaries

3 Control circuits for the circuit breakers and other equipment operated from the control room of the plant

4 Lighting circuits for the illumination of the plant and to provide power for portable equipment required in the upkeep and maintenance of the plant Sometimes special circuits are installed to supply the portable power equipment

5 Excitation circuits, which are so installed that they will receive good physical and electrical protection because reliable excitation is necessary for the operation of the plant

6 Instrument and relay circuits to provide values of voltage, current, kilowatts, reactive kilovolt amperes, and temperatures, and to serve the protective relays

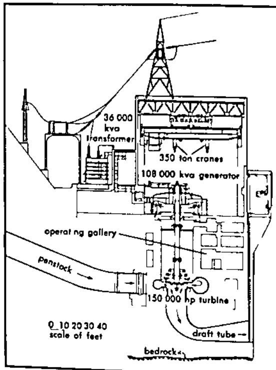


Fig 2 Typical layout and apparatus arrangement in a hydroelectric generating plant Cross section is through powerhouse and dam at a main generator position in the Grand Coulee plant

7 Communication circuits for both plant and system communications Telephone radio transmission line carrier and microwave radio may be involved

It is important that reliable power service be provided for the plant itself and for this reason station service is usually supplied from two or more sources

Generator protection Necessary devices are installed to prevent or minimize other damage in cases of equipment failure Differential current and ground relays detect failure of insulation which may be due to deterioration or accidental overvoltage Overcurrent relays detect overload currents that may lead to excessive heating, overvoltage relays prevent insulation damage Loss of excitation relays may be used to warn operators of low excitation or to prevent pulling out of synchronism Bearing and winding overheating may be detected by relays actuated by resistance devices or thermocouples Overspeed and lubrication failure may also be detected

Not all of these devices are used on small units or in every plant The generator

change in voltage for specific change in load (usually from full load to no load) expressed as percentage of normal rated voltage The voltage of an electric generator varies with the load, consequently some form of regulating equipment is required to maintain a reasonably constant and predetermined potential at the distribution stations or load centers Since the inherent regulation of most alternating current generators is rather poor (that is high percentagewise), it is necessary to provide automatic voltage control The rotating or magnetic amplifiers and voltage-sensitive circuits of the automatic regulators together with the exciters are all specially designed to respond quickly to changes in the alternator voltage and to make the necessary changes in the main exciter output thus providing the required adjustments in voltage A properly designed automatic regulator acts rapidly, so that it is possible to maintain desired voltage with a rapidly fluctuating load without causing more than a momentary change in voltage even when heavy loads are thrown on or off

Electronic voltage control has been adapted to some generator and synchronous condenser installations Its main advantages are its speed of operation and its sensitivity to small voltage variations

REGULATOR

Synchronization of generators Synchronization of a generator to a power system is the act of matching over an appreciable period of time the instantaneous voltage of an alternating-current generator (incoming source) to the instantaneous voltage of a power system of one or more other generators (running source), then connecting them together In order to accomplish this ideally the following conditions must be met

1 The effective voltage of the incoming generator must be substantially the same as that of the system

2 In relation to each other the generator voltage and the system voltage should be essentially 180 degrees out of phase, however, in relation to the bus to which they are connected their voltages should be in phase

3 The frequency of the incoming machine must be near that of the running system

4 The voltage wave shapes should be similar

5 The phase sequence of the incoming polyphase machine must be the same as that of the system

Synchronizing of ac generators can be done manually or automatically In manual synchronizing an operator controls the incoming generator while observing synchronizing lamps or meters and a synchroscope, or both Voltage (potential) transformers may be used to provide voltages at lamp and instrument ratings Lamps properly connected between the two sources are continuously dark when voltage phase and frequency are properly matched Wave shape and phase sequence are determined by machine design and rotation or terminal sequence Large units generally are provided with voltmeters and frequency meters for matching these quantities, and a synchroscope connected to both sources to indicate phase relationship Lamps may also be included The standard synchroscope needle

revolves counterclockwise when the incoming ma ch ne is slow and clockwise when fast The needle points straight up when the two sources are in pha e The operator closes the connecting switch or circuit breaker as the synchroscope needle slowly approaches the in phase position For discussion of the synchroscope *see* PHASE ANGLE MEASURE MENT PHASE METER

Automatic synchronizing provides for automati cally closing the breaker to connect the incoming machine to the system after the operator has properly adjusted voltage (field current) frequency (speed) and phasing (by lamps or synchroscope) A fully automatic synchronizer will initiate speed changes as required and may also balance voltages as required then close the breaker at the proper t me all without attention of the operator Auto matic synchronizers can be used in unattended sta tions or in automatic control systems where units may be started synchronized and loaded on a single operator command *See* ALTERNATING CURRENT GENERATOR ELECTRIC POWER SYSTEMS [E C S]

Electric power measurement

The measurement of the time rate at which work is done in an electric circuit The work done in moving an electric charge is proportional to the charge and the voltage drop through which it moves Charge per unit time defines electric cur rent electric power P is therefore defined as the product of the current I in a circuit and the voltage E across its terminals at a given instant Expressed symbolically $P - EI$

A second important definition of power follows d rectly from Ohm s law $P - I^2R$ where R is the resistance of the circuit

The practical unit of electric power is the watt The watt represents a rate of expending energy and thus it is related to all other units of power for example in mechanics 1 watt = 10^7 ergs/sec and 746 watts = 1 horsepower Commonly used small units are the milliwatt (0 001 watt) and the m crowatt (0 000001 watt) Large units are the k lowatt (1000 watts) and the megawatt (1 000 000 watts)

Power measurements must cover the frequency spectrum from direct current through the conven t onal power frequencies the audio and the lower radio frequencies to the highest frequencies (up to 25 000 kilomegacycles) In general different tech n ques are required in each frequency range and th s article is divided into sections dealing with the e frequency ranges *See* ELECTRICAL MEASURE MENTS

DC AND AC POWER FREQUENCIES

In the measurement of power in a direct current (dc) circuit not subject to rapid fluctuations there is usually no difficulty in making simultaneous observations of the true values of voltage and cur rent using common types of dc voltmeters and ammeters The product of these observations then gives a sufficiently accurate measure of power in the given circuit except that if great accuracy is re quired allowance must be made for the power used by the instruments themselves

If in a circuit the voltage e or current i or both are subject to rapid variations instantaneous values of power are difficult to measure and are usually of no interest The important value is the average value which may be expressed mathematically as

$$P = \frac{1}{T}\int_0^T ei\, dt$$

where T is the period or time interval and t is time This relation holds true for any waveform of cur rent and voltage In circuits with rapidly varying direct currents pulsating rectified current or in general alternating currents the continuous aver aging over short periods of time and the automatic multiplication of current and voltage values is ac complished by the wattmeter *See* WATTMETER

In ac circuits with steady effective values of volt age and current the voltmeter ammeter method may

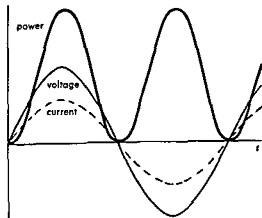


Fig 1 Curves of instantaneous current voltage and power in an ac c rcu t current and voltage n phase

be used as in the dc case except that of course ac meters are used and a phase meter is also required to measure phase angle unless current and voltage are in phase (*see* PHASE ANGLE MEASUREMENT) Because ac ammeters and voltmeters actually meas ure root mean square or effective values these lead directly to values of average power *See* AM METER VOLTMETER

Sinusoidal ac waves Figure 1 illustrates the case of a sinusoidal voltage and current in a circuit containing only a resistive load Here the current wave is entirely symmetrical with the voltage wave and the power curve formed from the product of the voltage and current at each instant appears as a double frequency wave on the positive side of the zero axis

ever would produ e a p

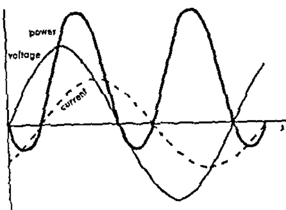


Fig 2 Curves of instantaneous current voltage and power in an ac circuit current and voltage out of phase

quadrature condition in which the current wave leads the voltage wave In general circuits contain elements of resistance inductance and capacitance in varying amounts and they must therefore assume some intermediate condition of phase angle between voltage and current In this case *the power* wave as shown in Fig 2 dips below the zero line and becomes negative indicating that during that part of the cycle power feeds back into the circuit Measurements based on readings of conventional ac ammeters and voltmeters do not account for these negative excursions and therefore in general the product of steady effective voltage and current readings in an ac circuit differs from and is *greater* than the reading of a wattmeter in such a circuit

In symbols these relationships in the *general ac* circuit may be expressed as follows

$$i = I_m \int_0^T \sin 2\pi ft \, dt$$

$$e = E_m \int_0^T \sin (2\pi ft \pm \phi) \, dt$$

where I_m and E_m are maximum values of current and *voltage* f is frequency in cycles per second and ϕ is the *phase angle by which* the current leads (+) or lags behind (−) the voltage in the circuit

But by definition

$$P = \frac{1}{T} \int_0^T ei \, dt$$

*Sub*stituting and carrying out the indicated operations

$$P = EI \cos \phi$$

where E and I are effective values of voltage and current

The above expression for P *is the* real or active power in the circuit and is distinguished from the simple product EI *which is* called the apparent or virtual power, *by the* factor $\cos \phi$, which is called the power factor It is obvious from the previous formula that

$$\cos \phi = P/EI$$

or that

$$\text{Power factor} = \frac{\text{real power}}{\text{apparent power}}$$

Negative or reactive power due to inductance and capacitance in a circuit, is given by the relation $EI \sin \phi$

The units for these quantities are for real power watts, for apparent power volt amperes, and for reactive power reactive volt amperes or vars For further discussion of real and apparent power in sine wave circuits *see* ALTERNATING-CURRENT CIRCUIT THEORY For a discussion of power in non sine-wave circuits *see* WAVEFORM NONSINUSOIDAL For the measurement of apparent and reactive power *see* VOLT AMPERES

Polyphase power measurement Summation of power in the separate phases of a polyphase circuit is accomplished by combinations of single-phase wattmeters, or wattmeter elements disposed according to the general rule called Blondel s theorem, as *follows* *If* energy is supplied to any system of conductors through N wires the total power in the system is given by the algebraic sum of N wattmeters so arranged that each of the N wires contains one current coil the corresponding potential coil being connected between that wire and some point on the system which is common to all the

no nt is on one of

attach-

measure-

ment may be effected by the use of $N - 1$ wattmeters

Considering measurement in three phase circuits as an example of polyphase practice in wide application in the power industry the two most common systems are the three-phase three-wire system in which the source may be Y connected or delta connected to three load wires or the three phase four wire system which also has three load wires and in addition a fourth wire or neutral which may or may not carry current to the *load* If the neutral does not carry current, the circuit may be treated as a three-wire system

Before applying the rule to commonly used circuits an exception may be noted in the case of a balanced circuit in which the effective *values of the* currents and voltages and the phase *relationships* between them remain constant In other words the loads on the separate phases are equal In *this* special case power may be measured by a single wattmeter connected in one phase and the reading multiplied by three

To measure power in either three wire or *four* wire systems a wattmeter may be connected in each of the power receiving circuits as in *Fig* 3 The sum of the three readings gives the total power

Alternatively in a three wire system total power may be measured by two wattmeters each having its current coil connected in one of the line conductors and its potential circuit connected between the line conductor in which its current coil is connected and the third line conductor (Fig 4) The algebraic

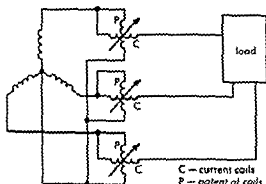


Fig 3 Three wattmeters in three phase three-wire circuit

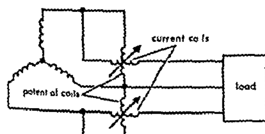


Fig 4 Two wattmeters in three-phase three wire circuit

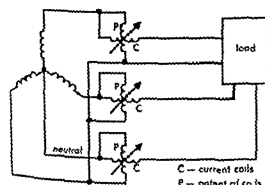


Fig 5 Three wattmeters in three-phase four wire circuit

sum of the readings of the two wattmeters indicates the total power in the three power receiving circuits

In a four wire system three wattmeters may also be effectively used by connecting the current coils in each of two of the line conductors and in the neutral conductor as in Fig 5 The potential coils are connected between each of the line conductors and the neutral conductor in which the respective current coils are connected and the third line conductor

In the three latter cases the methods are correct for any value of balanced or unbalanced load and for any value of power factor

A variety of other circuit connections is available for polyphase power measurement for various special conditions of use For further discussion of power in polyphase circuits *see* ALTERNATING-CURRENT CIRCUIT THEORY [C A M]

AUDIO AND RADIO FREQUENCIES

At frequencies above those used in the ordinary power distribution systems, dynamometer type wattmeters become inaccurate For measurements of transmitted power at audio and the lower radio frequencies no generally satisfactory substitute has been developed and measurements are therefore confined to determinations of power dissipated in a load or available from a source and are deduced from measurements of impedance and current or voltage

Power-output meters Power-output meters combine resistive loads which can be adjusted to various known values, and voltmeters calibrated to indicate the power dissipated therein They are used at audio frequencies to determine the maximum power output that can be obtained from a source or to measure output power in studies of harmonic distortion intermodulation overload frequency characteristic and so forth

Figure 6 shows an illustrative schematic A voltmeter V is fed from an attenuator having a constant input resistance R and a voltage ratio k The voltage V at the meter corresponds to a voltage kV at the attenuator input and the meter scale is calibrated in power input to the attenuator $P =$

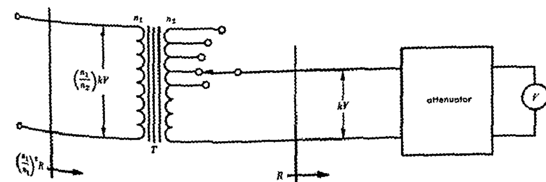


Fig 6 Schematic representation of audio-frequency power-output meter

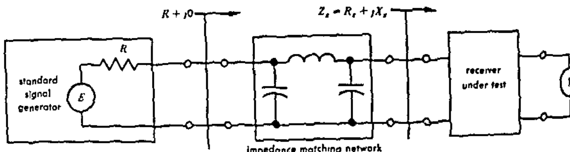


Fig 7 Circuit schematic showing method of using a standard signal generator and impedance-matching network to produce a known power input to a receiver

$(kV)^2/R$ By adjusting the attenuator one can obtain convenient scale multiplying factors k The tapped transformer T provides different turns ratios n_1/n_2 to adjust the effective input resistance of the instrument over a range of values

By adjusting this ratio the user can obtain a maximum meter reading when the corresponding input impedance is approximately equal to the output impedance of the source If the output impedance is purely resistive, this condition occurs when the impedances are exactly equal or matched, and the power indicated is the maximum power output that can be obtained from the source

Standard signal generators At radio frequencies the principle of impedance matching is frequently used to determine the power input to a receiver with a standard signal generator serving as source The circuit of Fig 7 shows the arrangement

A standard signal generator produces a known voltage E 'behind' an output resistance R Impedances at radio frequencies frequently have significant reactive components, thus the impedance-matching network serves the dual function of tuning out the reactive component X_r of the receiver input impedance Z_r and multiplying the resistive component R_r by a factor that depends upon the values of the component inductance and capacitances When the matching network is so adjusted that the receiver output is a maximum, the reactive component is nullified and the input resistance of the matching network equals the output resistance R of the standard signal generator The power input is then $P = E^2/4R$

Calorimetric method The power dissipated in a resistive load can be determined from the heat developed therein and several ingenious schemes for measuring this heat have been derived

A widely used technique depends upon dissipating the power in a resistive element of relatively high temperature coefficient of resistivity The power is deduced from the change in resistance by one of several methods Measurements of this general nature are known as bolometric methods See BOLOMETER

Another method depends upon using as the load an incandescent lamp that will emit visible light at power inputs of the order of those to be measured The temperature of the filament can be determined by pyrometric techniques, or the light can be measured with a photocell The indicating system can then be calibrated in terms of power input at direct current or low frequency alternating current, where accurate moving coil wattmeters can be used. Such calibrations will continue to hold good as the frequency of the source is raised until dielectric losses in the glass and eddy current losses in the metal parts other than the filament become significant compared with the heating of the filament proper, or until the wavelength becomes so short that the current flow in the filament becomes non uniform along its length Reasonable accuracy can be attained at frequencies as high as a few hundred megacycles

A variant of this method uses a diode in which a pure-metal filament is heated by the power to be measured When not space-charge limited, the dc plate current is a sensitive indicator of filament temperature The temperature rise in a heater element can also be measured by thermocouples See THERMOCOUPLE

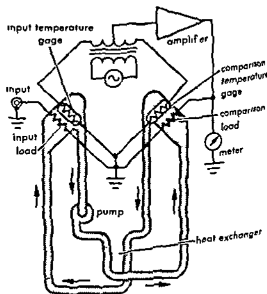


Fig 8 Commercial self balancing calorimeter for power measurements

The classical calorimetric method depends upon measuring the temperature rise in a liquid coolant circulated through an enclosure containing the load and so insulated that heat losses other than to the coolant are negligible *See* CALORIMETRY

Measurements of rate of temperature rise can be used for absolute determination of the power input in terms of the specific heat and volume of the liquid or the liquid can be used as a transfer mechanism to compare the radio-frequency power input with a known dc or low frequency power developed elsewhere in the system

Figure 8 shows a commercial instrument which uses a self balancing bridge to provide a comparison temperature rise that can be accurately calibrated in terms of a low frequency source

MICROWAVE FREQUENCIES

Measurement of dissipated power at microwave frequencies can be carried out by methods similar to those used at lower frequencies Because voltage and impedance become increasingly difficult to define precisely as the frequency is increased however calorimetric methods generally and bolometric methods specifically are almost universally used

In contrast to the lower radio frequencies convenient methods of measuring transmitted power are available at microwave frequencies These use the properties of waves propagated along transmission lines A commonly used device for this purpose is the directional coupler a simple example of which is shown in Fig 9 This comprises an auxiliary line coupled at two points spaced ¼ wavelength apart to the main line down which power flows For the sake of simplicity no coupling means is shown between the main and auxiliary lines other than holes in the outer conductors in practice either capacitive probes or small coupling loops may be used to transfer energy from one line to the other For the flow of power illustrated a wave entering the right hand hole and propagating to the left in the auxiliary line arrives at the left hand hole 180° out of phase with the wave entering the left hand hole If there are no losses in the system and the couplings through the two holes are equal the two signals cancel, and no signal reaches the left hand termination On the other hand a wave entering the left hand hole and propagating to the right in the auxiliary line arrives at the right hand hole in phase with the wave entering the right hand hole and the two signals add and propagate to the right in the auxiliary line Terminating the auxiliary line in a resistance equal to the characteristic impedance Z_0 prevents any of this power from reflecting in the auxiliary line and reaching the left hand termination Power traveling to the right in the main line therefore produces a signal voltage only at the right hand auxiliary line termination Conversely a signal propagating to the left in the main line reaches only the left hand termination If the two terminations are used as bolometers for instance they can then measure the individual powers traveling each way in the main line and the net power reaching the load which is equal to the difference between them *See* DIRECTIONAL COUPLER TRANSMISSION LINES

This measurement can also be made with a standing wave detector The voltage maximum is equal to the sum of the voltages of the incident and reflected waves $V_I + V_R$ whereas the voltage minimum is equal to the difference $V_I - V_R$ The power transmitted is the difference in power

$$\frac{(V_I)^2}{Z_0} - \frac{(V_R)^2}{Z_0}$$

which is equal to $V_{max}V_{min}/Z_0$ The relation between the probe voltage and the line voltage can be established by measuring power in matched loads In contrast to methods using directional couplers this has the advantage of requiring substantially no power however it requires manipulation to obtain a measurement *See* STANDING-WAVE DETECTOR [D B S]

Bibliography H Buckingham and E M Price *Principles of Electrical Measurements* 1957 E L Ginzton *Microwave Measurements* 1957 F K Harris *Electrical Measurements* 1952 F A Laws *Electrical Measurements* 2d ed 1938 C G Montgomery *Technique of Microwave Measurements* vol 2 1948 F E Terman and J M Pettit *Electronic Measurements* 2d ed 1952 M Wind (ed) *Handbook of Electronic Measurements* 2 vols 1958 M Wind and A Rapaport (eds) *Handbook of Microwave Measurements*, 2 vols 2d ed 1958

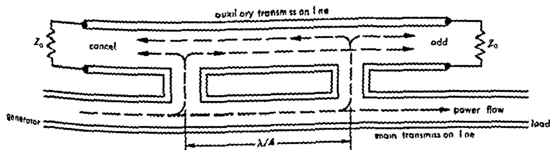


Fig 9 Cross sectional view of simplified two-hole directional coupler

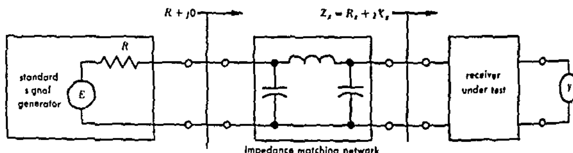


Fig 7 Circuit schematic showing method of using a standard signal generator and impedance matching network to produce a known power input to a receiver

$(kV)^2/R$ By adjusting the attenuator one can obtain convenient scale multiplying factors k^2 The tapped transformer T provides different turns ratios n_1/n_2 to adjust the effective input resistance of the instrument over a range of values

By adjusting this ratio the user can obtain a maximum meter reading when the corresponding input impedance is approximately equal to the output impedance of the source If the output impedance is purely resistive, this condition occurs when the impedances are exactly equal, or matched, and the power indicated is the maximum power output that can be obtained from the source

Standard signal generators. At radio frequencies the principle of impedance matching is frequently used to determine the power input to a receiver with a standard signal generator serving as source The circuit of Fig 7 shows the arrangement

A standard signal generator produces a known voltage E "behind" an output resistance R Impedances at radio frequencies frequently have significant reactive components, thus the impedance matching network serves the dual function of tuning out the reactive component X_x of the receiver input impedance Z_x and multiplying the resistive component R_x by a factor that depends upon the values of the component inductance and capacitances When the matching network is so adjusted that the receiver output is a maximum, the *reactive* component is nullified and the input resistance of the matching network equals the output resistance R of the standard signal generator The power input is then $P = E^2/4R$

Calorimetric method. The power dissipated in a resistive load can be determined from the heat developed therein and several ingenious schemes [illegible]

[illegible] power is deduced from the change in resistance by one of several methods Measurements of this general nature are known as bolometric methods See BOLOMETER

Another method depends upon using as the load an incandescent lamp that will emit visible light at power inputs of the order of those to be meas[illegible] The temperature of the filament can be deter[illegible] can then be calibrated in terms of power input at direct current or low frequency alternating current, where accurate moving coil wattmeters can be used Such calibrations will continue to hold good as the frequency of the source is raised until dielectric losses in the glass and eddy-current losses in the metal parts other than the filament become significant compared with the heating of the filament proper, or until the wavelength becomes so short that the current flow in the filament becomes nonuniform along its length Reasonable accuracy can be attained at frequencies as high as a few hundred megacycles

A variant of this method uses a diode in which a pure-metal filament is heated by the power to be measured When not space-charge limited, the dc plate current is a sensitive indicator of filament temperature The temperature rise in a heater element can also be measured by thermocouples See THERMOCOUPLE

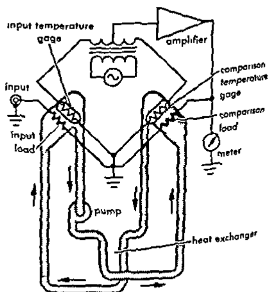


Fig 8 Commercial self balancing calorimeter for power measurements

For example *see* AMMETER, PHASE METER, VOLTMETER, WATTMETER

Control switchboards Control switchboards consist of one or more panels on which are mounted electrical devices to control and monitor the operation of remote electrical apparatus such as circuit breakers, generators, or motors They differ from a power switchboard in that main circuit switching or interrupting devices or their connections are not included

System control switchboards are attended by operators who observe the performance of the system and initiate changes by means of the controls provided

Devices commonly mounted on control switchboards include control switches rheostats protective relays control relays indicating and recording meters These devices are grouped physically in a manner convenient for operating and maintenance personnel The switchboard includes wiring to interconnect the devices mounted thereon Wiring is also required from the switchboard to the apparatus being controlled thus limiting the physical separation to a few thousand feet

Many substations and their control switchboards are unattended because of the development of automatic control and remote control systems which can include supervisory control

Automatic control Electric controls are arranged to provide for opening or closing or both in an automatic sequence and under predetermined conditions The required character of service is maintained and adequate protection is provided against all usual operating emergencies

Automatic operation, without the continuous presence of an operator, was first used on synchronous converter substations supplying an interurban electric railway in 1914 In the early 1920s the development of automatic reclosing relays protective relays and automatic voltage regulators enabled ac substations to be made completely automatic Automatic control has been applied to larger and larger substations and to many hydroelectric generating stations It is also used on individual circuits and equipment in many attended stations For example even in many stations where an operator is on duty outgoing overhead feeders are controlled by automatic reclosing relays

Supervisory control Under certain conditions it is impractical to make a station completely automatic The control equipment may be entirely too complex or it may be impossible to foresee all possible operating contingencies and arrange for suitable control In this case human judgment or knowledge of system conditions is a necessary prerequisite to proper control Such a circumstance however need not demand attendance at the station Control can be managed by means of supervision from some attended station

Supervisory control is a system for the selective control and automatic indication of remotely located equipment By means of supervisory control an operator at some distant point can perform such operations as the opening and closing of breakers or the starting and stopping of synchronous condensers or hydroelectric generators and he can receive an indication that the operation has been completed—all over a single channel that is a single pair of wires It is this use of a single common two way channel that distinguishes supervisory control from direct remote control, the latter requires one channel for each controlled unit

Supervisory control supplements rather than replaces fully automatic control All the protective relays and most if not all of the control relays that are used in completely automatic control of a device machine or circuit are equally necessary when the control is supervisory

The controlled station is known as the outlying station and the location from which it is controlled is known as the master station The master station can control more than one outlying station

A two way channel is needed between the master station and each outlying station so that signals can be sent both ways Using the outgoing signals from the master station the attendant effects a certain operation at the outlying station The automatic incoming signals indicate that a certain operation has occurred at the outlying station The following channels may be used (1) leased pair of wires (2) pair of wires in privately owned aerial or underground cable (3) privately owned open wire pair (4) carrier channel on privately owned open wire telephone pair (5) carrier channel on power line and (6) microwave

Protective relay systems Protective relays are designed to remove from service any element of a power system that suffers a short circuit or starts to operate in any abnormal manner that might cause damage or otherwise interfere with the effective operation of the rest of the system Protective relays accomplish this by tripping circuit breakers that are capable of disconnecting the faulty element

Circuit breakers controlled by relays are gen

cuit breakers are not economically justifiable fuses are used

Although the principal function of protective relaying is to mitigate the effects of short circuits it may also be used to detect other abnormal operating conditions for example thermal overload or reversed phase sequence of motors and generators A secondary function of protective relaying is to provide indication of the location and type of failure

All relays used for short circuit protection are operated from current and potential transformers connected in various combinations to the system element that is to be protected Through individual or relative changes in these two quantities protective relays can detect the presence type and ... tion of the failure For every type and ...

failure there is some distinctive difference in these quantities and there are various types of protective relays available each of which is designed to recog nize a particular difference and to operate in response to it Differences in each quantity are pos sible in one or more of the following (1) magni tude (2) frequency (3) phase angle (4) duration (5) rate of change (6) direction or order of change and (7) harm ›nics or wave shape *See* ELECTRIC PROTECTIVE DEVICES

Grounding of electric power systems A system is grounded if at least one point is intentionally connected to ground the grounded connection be ing either the neutral of a transformer or a genera tor The grounding connection may I e either direct or through a current limiting device The term electric power system is used here to designate a three phase combination of lines and associated apparatus connected together and operating at ap proximately the same voltage level (without inter vening transformers)

System neutrals are grounded primarily for these reasons (1) to limit overvoltages caused by neu tral to ground and line to ground faults to permit the application of minimum rated arresters to obtain the ultimate in surge protection for the connected apparatus and (2) to permit sufficient current to flow when the system becomes accidentally grounded to actuate protective relays to initiate the clearing of the trouble

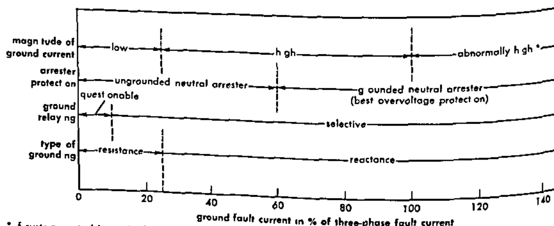
Electric power systems are operated either ungrounded or grounded An ungrounded or iso lated neutral system is a three phase system in which all the neutrals have no intentional connec tion to ground except possibly through indicating measuring or protective devices of very high im pedance A grounded neutral system is a three phase system in which some or all the neutrals are connected to ground either solidly or through re sistances or reactances of low value Effectiveness of the neutral grounding lies between the two extremes of the ungrounded system and the solidly grounded system In between these limits varying degrees of grounding may be obtained by connect ing the neutral point to ground through either a resistor or a reactor or a combination of both

A measure of the effectiveness of the groundi is the coefficient of grounding This is the rat expressed as a percentage of the rms voltage ground on a sound phase at the selected locati (generally the point of installation of equipmen resulting from system faults involving ground the line to-line voltage at the selected locati assuming the fault did not exist The ratio shall determined at a time during which the general subtransient reactances are effective If the coe cient of grounding is less than 80% the system said to be effectively grounded For an i olate neutral (ungrounded) system the coefficient 100%

Some of the basic factors which influence t selection of the coefficient of grounding a whether resistance reactance or no intentio neutral impedance is to be used are (1) sen itiv and selectivity of the ground relaying (2) requir degree of lightning arrester protection (3) possil mechanical and thermal damage due to h ground fault currents and (4) possible troub that may occur from transient overvoltages

Each of these factors must be weighed in sele ing the method of grounding and once the p dominant one is decided upon it is possible evaluate the influence of the other factors A su mary of conditions under which resistance a reactance grounding may be applied in terms ground fault current magnitude arrester prot tion and ground relaying is given in Fig 3

Lightning protection Substations normally protected against lightning because the expense protection is less than the repair of damag equipment Protection is also required to assure desired degree of continuous electrical supply the load being served For a general discussion LIGHTNING AND SURGE PROTECTION



* f system neutral is establ shed through generator a ground ng impedor is re qu red to reduce generator ground current contr but on to less than 100% of generator three phase fault current

Fig 3 Chart s gu de to res stance and reactance appl cat on for system ground ng

Protection against direct strokes is most often provided by lightning rods. Figures 1 and 2 show steel pipe used to extend the columns of the structure to the desired height so that satisfactory cones of protection are achieved.

To avoid direct strokes of lightning to a line immediately adjacent to the substation, overhead circuits often have overhead ground wires for the first half mile or so adjacent to the substation, regardless of whether the remainder of the line is shielded. The uppermost wires in Fig. 1 are ground wires. Substations can also be shielded by extensions of these overhead ground wires over the station.

Protection against traveling waves of voltage or surges entering the station due to more remote lightning strokes, or against switching surge voltages caused by the operation of circuit interrupting devices, is usually provided by lightning arresters. Since this protection decreases with physical separation, a set of arresters usually is located physically adjacent to transformers. When circuit breakers are closed, they receive some protection from these arresters and additional help from the surge dividing action where several circuits are connected to the station. However, a circuit breaker may be tripped open during a lightning storm. Complete protection for breakers under this condition is provided by additional lightning arresters located on the line side of the circuit breakers; some installations accept the lesser protection provided by the use of line side gaps or the increased hazard of no additional line side protection. [J.A.S.M.]

Bibliography: American Institute of Electrical Engineers (AIEE), *Standards for Neutral Grounding Devices*, AIEE Standard 32, 1947; American Standards Association, *American Standard Definitions of Electrical Terms*, ASA C42.35, 2d ed., 1957; Substation one-line diagrams, *Trans. AIEE*, 72(3):747–752, 1953.

Electric power systems

An assemblage of equipment to generate, transmit, and distribute electric energy. The main elements of a power system are shown in Fig. 1. They are outlined briefly here and discussed more fully in the cross-referenced articles.

Generation. The source of electric power is one or more generating stations or power plants. The generating station is an energy conversion unit. The energy source is a fuel (coal, gas, oil), water power, or nuclear reaction, and the energy output is electricity. *See* ELECTRIC POWER GENERATION.

More than 75% of the electric energy generated in the United States is obtained from generators driven by steam prime movers (mostly steam turbines). The largest steam turbine generator is rated at 500,000 kilowatts (kw).

Hydroelectric stations contribute most of the rest of the total power supply. In hydro plants water wheels (hydraulic turbines) drive the generators. The largest projected unit (1958) is rated 150,300 kw, requiring a 200,000-hp turbine. A single project will use 13 of these hydro units.

Internal combustion engines drive generators in many small power plants, but their total output is a small part of the total power capacity of the country. The largest of these generators is about 10,000 kw.

Gas turbine driven generators are installed in a few power plants in the United States. Typical ratings are below 7500 kw, although a 22,000-kw unit is planned for supplying power for short periods when electrical loads are unusually high. This application, if successful, may initiate a new trend in power generation.

Nuclear power stations for commercial service are in their infancy. One in operation at Shippingport, Pennsylvania, is rated 60,000 kw. Another, the Calder Hall plant in England, has a capacity of 150,000 kw in eight electric generating units supplied by four reactors of 205,000 kw thermal capability.

Because of their simplicity and efficient use of conductors, alternating current (ac), three-phase, 60-cycle systems are used almost exclusively for supplying electric power in the United States. Therefore, power system generators are of this type. The output voltage of generators is limited by design features and is usually in the range of 11,000–20,000 volts. These generator voltages are stepped up by transformers, usually at the generating station, for transmission of power to distant areas.

Transmission. The transmission system delivers electric power from the generating stations to load areas efficiently and in large amounts. They are designed to allow interconnection with neighboring systems to effect economies by operating on an efficient regional basis.

Transmission circuits are designed in present practice for voltages of 110,000–345,000 volts. For a given distance of transmission, the permissible power loading varies as the square of the voltage. Therefore, high transmission voltages are desirable. The power capacity of a 300-mile 3-phase line for different voltages is given in Table 1. The increase in transmission levels from 1890–1960 is shown in Fig. 2. These high voltages are stepped down by transformers in bulk power substations to levels of 23,000–69,000 volts. At the lower voltages power is fed to many miles of subtransmission lines extending to area load centers. Large industrial users are often supplied at subtransmission voltages. *See* TRANSMISSION LINES.

Substations. Ordinarily, substations are designed to step down the high voltage to a lower voltage. On transmission systems, the substations

Table 1. Power loading of a 300-mile 3-phase transmission line

Line-to-line voltage, kv	Loading, kva
100	25,000
200	100,000
300	225,000

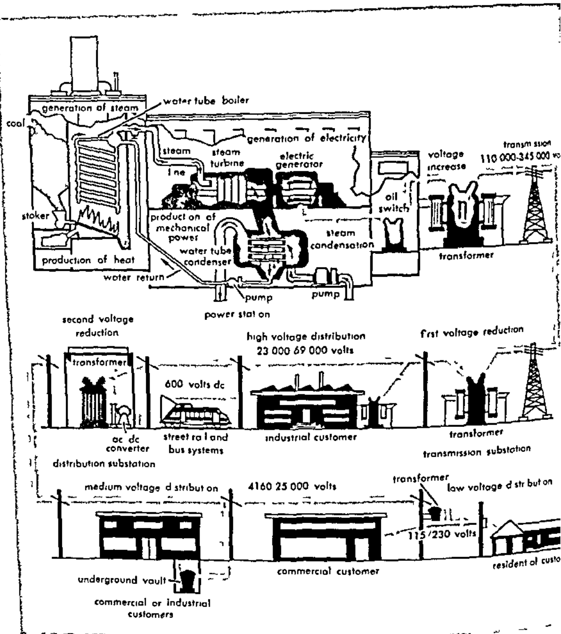


Fig 1 Major steps in the generation transmission and distribution of electricity

are generally termed bulk power substations on distribution systems they are known as distribution substations, on or near industrial customers premises they are designated industrial substations

Basic equipment in a substation includes transformers circuit breakers disconnect switches, and associated devices See ELECTRIC POWER SUBSTATION

Distribution A distribution system is that part of an electric power system between the bulk power source and the customers' switches The system includes primary lines distribution transformers secondary lines service drops (or lines) and associated circuit devices

Primary distribution circuits are usually operated at 4 160 to 25 000 volts These lines may supply large commercial institutional and some indu loads directly Smaller consumers are sup through numerous service transformers The cuits are carried on poles or are placed u ground

At conveniently located transformers volta stepped down to 115 and 230 volt levels fo secondary lines Service drops are run fro secondaries to the residential rural office small industrial plant loads These low voltag known as utilization voltages See ELECTRIC D BUTION SYSTEMS

Electric utility industry The electric util dustry is one of the largest industries in the U States, having an income close to $9 750 00 that is derived from revenues This industry

Table 2 Electric utility industry statistics (1948–1958)*

Item	1958	1956	1954	1952	1950	1948
Generating capacity† installed kw × 10³	142 413	120 697	102 592	82 226	68 919	56 560
Electric energy produced kwhr × 10⁶	644 760	600 668	471 686	399 224	329 141	282 698
Energy sales						
Total kwhr × 10⁶	569 161	530 128	410 904	342 524	280 539	240 740
Residential	159 047	133 851	108 165	86 780	67 030	50 978
Commercial	101 213	87 743	73 373	62 080	50 446	43 193
Industrial	275 023	276 647	200 155	167 358	139 065	124 088
Rural‡	11 084	11 098	10 176	8 536	7 100	6 327
Other	22 791	20 789	18 735	17 770	16 598	16 154
Customers × 10³	56 208	53 995	51 215	48 451	44 986	40 722
Revenue × $1 000 000	9 734	8 698	7 277	6 137	5 086	4 313
Average residential use kwhr	3 366	2 969	2 549	2 169	1 830	1 563
Average bill rate cents/kwhr	2.53	2.60	2.69	2.77	2.88	3.01
Coal consumption rate lb/kwhr	0.91	0.94	0.99	1.10	1.19	1.30
Average energy use kwhr/kw of average capacity	4 749	5 108	4 860	5 054	4 986	5 193

* From *Electrical World* Feb 23 1959
† Fuel and hydro
‡ Sales at rural rates

prises 3373 public and privately owned systems producing electricity in 3070 generating plants. Installed generating capacity is some 142 413 000 kw and annual electric production is over 644 760 000 000 kwhr (as of January 1 1959)

The 3373 systems comprise 461 private companies 1889 municipal systems 927 rural electric cooperatives 59 public power districts 10 irrigation districts, 9 U S government systems 9 state owned authorities 2 county systems and 7 mutual systems In *addition* in *U S* possessions there are 32 systems comprising 12 private companies 11 municipal systems 7 rural electric cooperatives one irrigation district and one state-owned system These systems operate 57 plants

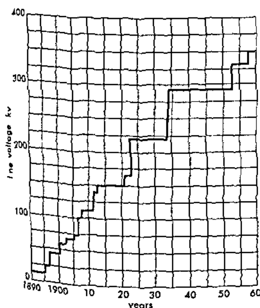


Fig 2 Growth of transmission voltages from 1890 to 1960

True proportions of the electric industry in the United States are best illustrated by Table 2 An idea of the industry s growth is indicated by the percentages (in parentheses) for a 10 year period from 1948 At the end of 1958 the industry s installed generating capacity reached 142 413 000 kw (a 125% increase over 1948) This capacity produced 644 760 000 000 kwhr (128% increase) resulting in sales of 569 161 000 000 kwhr (136% increase) to 56 210 000 customers (38% increase) Revenues from energy sales totaled $9.7 billion during 1958 (125% increase)

Residential use of electric energy set a new high of 3366 kwhr in 1958 (115% increase over 1948), and the residential average billing rate or charge fell to 2.53 cents per kwhr (about 16% less than in 1948) [R M SH]

Bibliography *Directory of Electric Utilities* R M Gardner TVA builds 500 Mw steam unit *Elec World* 151(20) 70–72 1959 A E Knowlton (3) 34–37 1959 A J Stegeman Electrical world industry statistics (with revision) *Elec World* 151(8) 83–89 1959 Study starts on gas turbines diesels *Elec World* 151(19) 38 1959 J H M Sykes Calder Hall completes first two years *Elec World* 151(13) 67 70 1959

Electric protective devices

A particular type of equipment applied to electric power systems to detect abnormal conditions and to initiate appropriate action to correct the abnormal condition From time to time disturbances in the normal operation of electric power systems occur These may be caused by natural phenomena such as lightning wind or snow, by accidental means traceable to reckless drivers inadvertent acts by plant maintenance personnel or other acts of

human beings or by conditions produced in the system itself such as switching surges or load swings Protective devices must therefore be installed on a power system to ensure continuity of electric service to prevent injury to personnel and to limit damage to equipment when abnormal situations develop

Protective devices like any type of insurance are applied commensurately with the degree of protection desired For this reason application of protective devices varies widely from installation to installation

This article first discusses protective relays as a basic device used in protective systems Six common abnormal c nditions for which protection is desired are then discussed For other areas of protection see GROUNDING ELECTRICAL LIGHTNING AND SURGE PROTECTION

Protective relays These are used to sense changes in the voltages and currents of a power system Sufficiently large variations from normal in these quantities cause the relay to operate Operation of the relay results in opening of circuit breakers to isolate that portion of the power system experiencing an abnormal voltage or current condition A fault in one part of the system affects all [illegible]

late equipment near the fault to prevent excessive damage or personal hazard

Relays are built to respond to voltage current or a combination of voltage and current Operation of the relay either opens or closes a contact Two basic principles are used in the construction of protective relays The simplest type of relay operates on the electromagnetic attraction principle This relay is composed of a coil plunger and set of contacts as illustrated in Fig 1 When current

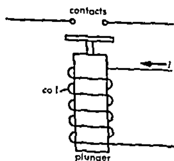


Fig 1 Plunger relay

flows in the coil a force is produced that causes the plunger to move and close the relay contacts

The electromagnetic induction principle is also used as a basic building block in construction of induction relays This type of relay responds to alternating current only whereas the relays discussed above respond to either direct or alternating current Briefly an induction relay consists of an electromagnetic circuit a disk or other form of rotor made of a nonmagnetic current-carrying material and contacts

A schematic illustration of an induction type relay is shown in Fig 2 The main coil is connected to an external source When current flows in the

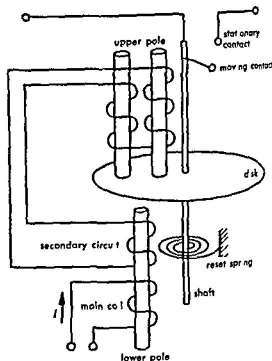


Fig 2 Induction relay

main coil transformer action induces current in the secondary circuit connected to the upper poles Fluxes produced by the currents flowing in the various pole circuits induce eddy currents in the rotor disk Interaction between rotor eddy currents and the fluxes passing through the rotor produces a torque on the rotor, causing it to move and thus closing the contacts This will be recognized as the split phase motor principle where two out of phase fluxes produce torque in a rotor A spring is used to reset the disk after the relay has operated *See* RELAY

By use of the principles of electromagnetic attraction and electromagnetic induction protective relays can be built to respond to all abnormal conditions that may occur in practice

Overcurrent protection This must be provided on all systems to prevent abnormally high current from overheating and causing mechanical stress on equipment Overcurrent in a power system usually indicates that current is being diverted from its normal path by a short circuit In low voltage distribution type circuits like those found in homes adequate overcurrent protection can be provided by fuses that melt when current exceeds a predetermined value (*see* FUSE ELECTRIC) Small thermal type circuit breakers also provide overcurrent protection for this class of circuit As the size of

circuits and systems increases the problems associated with interruption of large fault currents dictate the use of power circuit breakers Normally these breakers are not equipped with elements to sense fault conditions, and therefore overcurrent relays are applied to measure the current continuously When current reaches a predetermined value the relay contacts will close This actuates the trip circuit of a particular breaker, causing it to open and thus isolating the fault *See* CIRCUIT BREAKER

Either induction or plunger relays can be used to detect overcurrent condition As the current in either type of relay increases, the resultant force also increases When sufficient force is available the relay contacts close Relays have a well defined time-current characteristic, that is a longer time is required to close the contacts on a relay measuring a slight overcurrent A shorter time is required to close the contacts on a relay measuring a heavy overcurrent

Overvoltage protection This is usually applied on generators that are subject to overspeed when the load is lost A voltage which is higher than normal places a severe stress on insulation If the insulation should fail, a current path to ground would result and above normal current would then produce additional damage to the equipment Overvoltage relays are installed to detect this condition at locations where overvoltage conditions would be harmful Either induction or plunger relays can be set to trip appropriate circuit breakers at a predetermined value of voltage

Undervoltage protection This must be provided on circuits supplying power to motor loads Low voltage conditions cause motors to draw excessive currents which can damage the motors If a low voltage condition develops while the motor is running the relay senses this condition and removes the motor from service

Undervoltage relays can also be used effectively prior to starting large induction or synchronous motors These types of motors will not reach their rated speeds if started under a low voltage condition Relays can be applied to measure terminal voltage and if it is below a predetermined value the relay prevents starting of the motor

Underfrequency protection. This may be necessary in industrial plants where the load is supplied by a combination of local generators and a tie to an outside power company If the power company tie is disconnected local generators become overloaded and the frequency drops Underfrequency relays detect this condition and act to disconnect part of the load thereby preventing damage to the generators Underfrequency protection is also used to disconnect certain selected loads automatically or to sectionalize a transmission system when system frequency drops below a predetermined value Induction type relays are used for underfrequency protection

Reverse current protection This is provided when a change in the normal direction of current indicates an abnormal condition in the system Plunger relays applied to dc circuits sense a change in direction of current by a reversal of the direction of force on the plunger

Several schemes may be applied on ac circuits to sense a change in current direction, the most common can best be illustrated by an example Current at one voltage level flows into one side of a transformer and current at a different voltage level flows out the other side If a fault of some type occurs inside the transformer connected in a complex system a different condition then exists Current flows to the fault from both sides Simple overcurrent relays applied to the primary side of the transformer may sense the increase in current and disconnect the primary from the system but a low current from the secondary side may not be sufficiently large to make the secondary relay operate

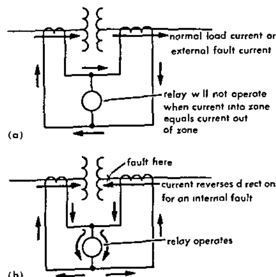


Fig 3 Simple differential relay scheme for reverse current protection (*a*) Normal load condition (*b*) Internal fault condition

Another scheme however immediately senses the reverse current in the secondary For a normal load condition current flows as shown in Fig 3*a* and the relay is unaffected When an internal fault occurs the current on the secondary of the transformer reverses direction Current then flows through the relay and causes it to operate thus [illegible] also [illegible] direction of rotation is important electric motors must be protected against phase reversal A reverse phase rotation relay is applied to sense the phase rotation This relay is a miniature three phase motor with the same desired direction of rotation as the motor it is protecting If the direction of rotation is correct, the relay will let the motor start

If it is incorrect the relay sensing this condition will prevent the motor starter from operating

[J M C]

Bibliography C R Mason *The Art and Science of Protective Relaying* 1956, Westinghouse Electric Corporation *Electrical Transmission and Distribution Reference Book* 4th ed 1950

Electric rotating machinery

Any form of apparatus which generates converts transforms or modifies electric power having a rotating member The most common forms are motors generators synchronous condensers synchronous converters rotating amplifiers phase modifiers and combinations of these in one machine The capacity or rating is usually indicated on a nameplate and denotes the maximum continuous duty without overheating or other injury Motors are rated in horsepower They are built in sizes from a small fraction of a horsepower to more than 100 000 hp Generators are rated in kilowatts or kilovolt amperes They range up to well over 300 000 kva Other types of rotating machines fall within these rating limits

Construction Most rotating machines consist of a stationary member called the stator and a rotating member called the rotor The rotor may be supported in bearings at both ends or it may be supported at one or both ends by the shaft of another machine

The illustration shows a typical rotating machine having a bracket bearing at one end and an arrangement for coupling to a turbine shaft at the other Although small machines sometimes employ *antifriction* bearings larger units are built with sleeve bearings generally lined with babbitt Vertical shaft machines use thrust bearings to support the rotating *member* Lubrication in slow or medium speed units is often supplied from an oil reservoir contained within the bearing housing Where bearing losses are high water cooling coils may be immersed in the oil to prevent overheating High speed machines are often lubricated from a pressurized oiling system, which also supplies the shaft seals in hydrogen cooled units

To function properly, rotating machines must have a magnetic circuit usually involving both rotor and stator, and one or more insulated electrical circuits which interlink the magnetic circuit To afford a low reluctance magnetic path the rotor and stator are separated only by a small clearance called the air gap

The windings are insulated electrically with materials such as enamel cotton varnished cambric mica asbestos dacron and glass fabric The most common impregnants are shellac asphaltum base varnish and epoxy or phenolic resins External partially conducting varnish is sometimes applied to high voltage coils for corona shielding

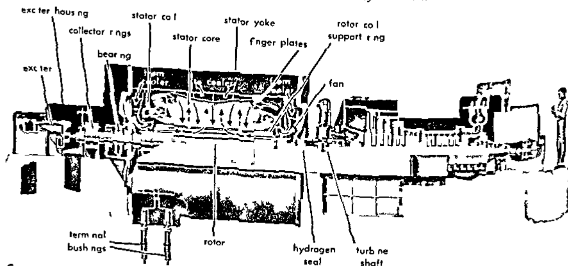
Principles of electrical mechanical energy conversion The force F in dynes produced on a conductor located at right angles to a magnetic field is

$$F = BIL/10 \quad \text{dynes}$$

where B is the flux density in lines per square centimeter in the vicinity of the conductor I is the conductor current in amperes and L is the length of the conductor exposed to the flux

In a motor the magnetic field created by one member exerts a force on the current-carrying conductors of the other, producing a *mechanical* torque which drives the load In a generator the changing magnetic field induces voltage in the armature windings when the rotor is driven by a source of mechanical power Little power is required at no load but as the load current builds up the prime mover must supply the torque to overcome the forces in the equation between the field and conductors

Ventilation Rotating machinery must be ventilated to avoid overheating from internal losses The cooling medium usually air or hydrogen is circulated by fans or blowers mounted on the rotor



Cross section of a typical electric rotating machine (Allis Chalmers)

or separately driven The illustration shows axial flow fans at each end of the rotor with arrows indicating the path taken by the gas With conventional cooling the cooling medium is blown over the exposed surfaces of the insulated windings and core With conductor cooling the cooling medium flows in ducts within the major insulation wall This is more effective and is used in the largest machines *See* ALTERNATING-CURRENT GENERATOR

Losses In all rotating machines losses occur Among them are I^2r losses called copper losses in the windings connections and brushes stray load losses in windings solid metal structures and frame core loss in the magnetic material and structural parts (*see* CORE LOSS) windage and friction loss, and exciter and rheostat losses

I^2r losses (in watts) in each path of the windings are equal to the square of the effective current in amperes times the resistance in ohms Brush I^2r loss is the product of the potential drop in volts times the current in amperes Stray load losses are caused mainly by eddy currents, due to variable magnetic fields (produced by the load current) within the conductors pole surface structural members end shields frame and so forth

Windage and friction losses These result from circulation and turbulence of the cooling medium and friction of bearings seals and brushes Windage loss is relatively large in air cooled high speed machines In hydrogen the loss is only 7 15% of that in air within the operating range of purity Bearing and seal friction losses are generally absorbed by the lubricating oil To avoid excessive friction or overheating of bearings an inlet oil temperature of 100–120°F is often recommended for large machines discharging at about 150°F [L T R]

Bibliography B F Bailey and J S Gault *Alternating current Machinery*, 1951, A E Knowlton (ed) *Standard Handbook for Electrical Engineers* 9th ed 1957 A F Puchstein T C Lloyd and A G Conrad *Alternating Current Machines* 3d ed 1954

Electrical codes

Systematic bodies of rules governing the practical application and installation of electrically operated equipments and devices and electric wiring systems The purpose of such codes is the practical safeguarding of persons and of buildings and their contents from hazards that may arise from the use of electricity for light heat power radio signal ing and for other purposes

National Electrical Code The basic electrical code in the United States is the National Electrical Code standard of the National Board of Fire Underwriters for electric wiring and apparatus as recommended by the National Fire Protection Association It is an approved American standard by the American Standards Association

The first nationally recommended electrical code was published by the National Board of Fire Underwriters in 1895 With this code as a basis the National Electrical Code was drawn up in 1897 the result of the united efforts of the various insurance electrical architectural and allied interests which through a conference composed of delegates from various national associations recommended it to these respective associations for adoption In 1911 the work of the conference was taken over by the National Fire Protection Association which acts as sponsor for the project under the rules of procedure of the American Standards Association

The provisions of the National Electrical Code are under constant review by a number of code making panels each consisting of a number of members selected to provide broad representation of electrical industry and public interests The Code is amended or added to (1) by general revisions incorporated in its periodic republication at intervals of 3–5 years or (2) by tentative interim amendments which are announced by bulletins and through the technical press

Municipal codes The National Electrical Code is incorporated bodily or by reference in many municipal building ordinances often with additional provisions or restrictions applicable in the particular locality Some large cities have independent electrical codes however the actual provisions in most part tend to be substantially similar to the National Electrical Code

Administration Electrical codes are administered locally by electrical inspectors who review plans and specifications and inspect electrical work during installation and after the work is completed to ensure compliance with applicable rules or ordinances

Electrical inspection bureaus are maintained in many cities by the National Board of Fire Underwriters In communities where codes are enforced by ordinance inspections may be performed by municipal electrical inspectors Utility inspectors inspect the service entrance and metering installation for compliance with prevailing utility regulations

Federal and state buildings are usually inspected by authorized federal or state electrical inspectors In these instances inspection includes both safety consideration and the requirements of the particular job specifications Other inspection (by underwriters or municipal inspectors) is often waived *See* WIRING (ELECTRIC) [W T S]

Bibliography A L Abbott and C L Smith *National Electrical Code Handbook* 9th ed 1958

Electrical conduction in gases

The process by means of which a net charge is transported through a gaseous medium It encompasses a variety of effects and modes of conduction ranging from the Townsend discharge at one extreme to the arc discharge at the other The current in these two cases ranges from a fraction of 1 microampere in the first to thousands of amperes in the second It covers a pressure range from less than 10^{-4} atm to greater than 1 atm *See* ARC DISCHARGE TOWNSEND DISCHARGE

process Not only does the gas permit the drift of free charges from one electrode to the other but the gas itself may be ionized to produce other charges which can interact with the electrodes to liberate additional charges Quite apparently the current voltage characteristic may be nonlinear and multivalued See ELECTRICAL CONDUCTIVITY OF METALS ELECTROLYTIC CONDUCTANCE SEMICONDUCTOR

The applications of the effects encountered in this area are of significant commercial and scientific value A few commercial applications are thyratrons gaseous rectifiers ignitrons glow tubes and gas filled phototubes These tubes are used in power supplies control circuits pulse production voltage regulators and heavy duty applications such as welders In addition there are gaseous conduction devices widely used in research problems Some of these are ion sources for mass spectrometers and nuclear accelerators ionization vacuum gages radiation detection and measurement instruments and thermonuclear devices for the production of power

The discussion of this complicated process will be divided into two parts The first will deal with the basic effects involved including production and loss of charges within the region and the motion of charges in the gas The second part will deal with the mechanism of conduction

BASIC EFFECTS

To produce gaseous conduction two conditions must obtain First there must be a source of free charges Second there must be an electric field to produce a directed motion of these charges Considering the first of these one finds that the free charge concentration is a result of a number of processes which produce and remove charges

Sources of free charges and ionization In many
ode is
from
discus
SION) Closely related to this as a source of electrons is field emission Here a strong positive field at a metallic surface lowers the barrier for electron emission Thus the electron current from a surface at a given temperature may be significantly increased See FIELD ENHANCED EMISSION

Both of these effects result in electron production Another effect photon absorption may result only in electrons if the absorber is a solid (see PHOTOELECTRICITY) However if the photon interacts with a gas molecule or atom ionization may result In this case, both an electron and a positive ion are obtained The photon may come from some external source or it may be a secondary effect of the gaseous conduction It may have a wavelength in the visible ultraviolet or x ray region

Conduction in flames is largely a result of the thermal production of ionization This is a specialized field which has long been of interest in chemistry and combustion studies To produce appreciable thermal ionization, the temperature must be high as in a flame If the effective temperature is known the ionization concentration may be determined from statistical mechanics Thermal ionization is also of tremendous importance in devices for production of power by thermonuclear processes and in ion propulsion equipment A special form of this is surface ionization in which a hot surface may cause ionization of a gas atom that comes in contact with it

Another source of ionization is particle radiation due to cosmic rays, radioactive material in the walls or in the gas or particles produced from an external source These particles may then produce an ionized track in the gas An example is the ionization produced by an α particle from a polonium source in an ionization chamber

In most of these methods of charge production the sources are primary That is the presence of other free charges is not important in the production of ionization or electrons Other processes are secondary in origin although they may be of prime importance It was pointed out that photons could originate either externally or as a secondary effect Field emission could be a secondary effect also Other methods of ionization however are generally thought of as being secondary in origin Ionization of the gas by electron impact is such a case Here free electrons may gain enough energy in an electric field to interact with an atomic or molecular electron to produce an ion pair

Cumulative ionization is an extension of ionization by impact If the original electron and its offspring gain enough energy so that each may produce another electron and if this process is repeated over and over the result is called an avalanche and the ionization thus produced is referred to as cumulative ionization This is the basis for particle detection in some ionization devices See GEIGER MULLER COUNTER, IONIZATION CHAMBER

Another secondary source is electron emission from either electron or positive ion bombardment of a surface (see SECONDARY EMISSION) This should not be confused with thermionic emission resulting from heating under bombardment

Other sources which may be important are atomic collisions sputtering and collisions of the second kind In the first case an atom or heavy ion may collide with an atom to produce ionization This is quite unlikely until an energy of many times the ionization energy is obtained The second is somewhat analogous to secondary electron emission Here the positive ions strike a surface and knock out atoms or groups of atoms Some of these come off as ions In the third case an excited atom may interact with an atom or molecule which is chemically different and has an ionization potential lower than the excitation potential of the excited

atom The result may be the decay of the excited state with ionization of the struck molecule or atom Symbolically



$$A^* + B \rightarrow A + B^+ + e$$

where A* is the excited atom B the struck atom or molecule and *e* an electron *See* EXCITATION PO TENTIAL IONIZATION POTENTIAL SPUTTERING

Free charge removal The net free charge con centration is a balance between charge-production and charge-removal processes Recombination is one s ich process Here an electron or heavy nega t e ion and a positive ion may recombine The energy transition may appear as electromagnetic radiation or may be carried off by a third body if one is present There are a wide variety of condi tions which may lead to recombination Where the temperature and electric field are high the recom b nation will occur predominantly at the walls

The method of charge removal is important from the aspect of conduction however If the charges move to the appropriate electrodes under the in fluence of the field and there recombine then they contribi te to the current If they simply diffuse to the walls and recombine there with ions of the opposite sign or if they recombine in the gas volume they may not appear as part of the ex ternal current

Motion of the charges The motion of the charges within the gas will be largely influenced by the potential function For the usual regular geometries this could be calculated in principle if there were no charges present However in a gas with free charges distributed throughout the prob lem is q ite different The charges modify the charge free potential but the potential itself deter m nes how charge will move The motion of the charges further modifies the potential and so on Although the situation can be described physically by Poiss n on

u ents which are made with probes This req iires caref l

do

sul ui thermal or agitation energy The random ness of the motion is brought about by the many collisions with molecules and other ions A great d fference exists in the motions of electrons and heavy ions Because of low mass electrons are ea ly deflected so they move erratically They d ffu e badly and follow field lines only generally Again because the mass is small an electron can g e up appreciable energy only in an inelastic col l on in which excitation or ionization takes place Hence electron agitation and diffusion will be m ch greater in a pure inert gas than in a gas ha ng many low energy molecular states Con versely heavy ions exchange energy effectively at every coll sion Diffusion is much less so that they follow the electric field lines more closely than do electrons

MECHANISM OF CONDUCTION

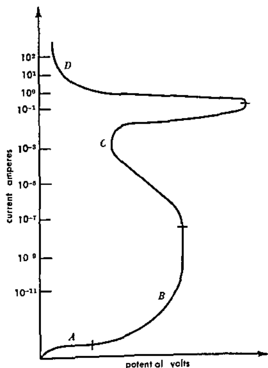
The ionic mobility μ relates drift velocity v to electric field X by the equation

$$v = \mu X$$

For electrons the mobility is high and a drift velocity of 10^6 cm/sec or greater may be obtained The electronic mobility is not a true constant but varies with field pressure temperature and gas composition For heavy ions the mobility is much more nearly constant but is still dependent on these quantities to some extent Drift velocities are usually of the order of 10^3 10^4 cm/sec Thus in a typical conduction device an electron may move from one electrode to the other before a heavy ion is displaced appreciably

It would appear from the foregoing that if accurate information about the important processes existed one could predict the characteristics of the conduction process under given conditions Un fortunately this is not the case Generally the situation is so complicated that the theory can yield only qualitative predictions Accordingly most of the information concerning the various forms of gaseous conduction is empirical In the present description it will be possible to mention the main features of a few of these modes

The illustration shows a sample voltage current characteristic for a two electrode device with con stant pressure It is assumed that there is a con stant source of ionization which could be any of



Current potent al character st cs for a two-electrode de vice w th constant pressure

the primary sources previously discussed In region A the current first rises and then over a limited range is relatively constant as the voltage across the electrodes is increased The initial rise is the result of the collection of charges which were either recombining or diffusing to the walls The nearly constant current region is the result of the collection of almost all of the charges

In region B further increase in voltage produces an increase in current Here ionization by electron impact is occurring The situation is described by specifying that each free electron makes α additional ion pairs in traveling 1 cm in the direction of the field The number of ion pairs produced per second in 1 cm at a distance x from the cathode (assuming parallel plate electrodes) is given by

$$n = n_0 e^{\alpha x}$$

where n_0 is a constant depending on the initial number of electrons This is a form of the Townsend equation and α is the first Townsend coefficient In the region B in the figure the increase in current represents an increase in α Near the end of this region the current increases more rapidly with applied field Here additional effects are taking place such as the photoelectric process and secondary emission This situation is described by

$$i = i_0 \frac{(\alpha - \beta) e^{(\alpha - \beta) x}}{\alpha - \beta e^{(\alpha - \beta) x}}$$

where β is the second Townsend coefficient i_0 is the initial electron current at the cathode and i is the anode current as a function of plate separation x, β is also a function of electric field

At the end of the region the slope becomes infinite and if the external resistance is not too large the current will jump in a discontinuous fashion The transition is referred to as a *spark* and the potential at which it occurs is the breakdown or sparking potential The region B is called a Townsend discharge and is not self sustained Thus if the source of primary ionization were removed the discharge would cease See BREAKDOWN POTENTIAL SPARK ELECTRIC

As the potential reaches the sparking potential a transition occurs to the region C This is the self sustained glow discharge region Over an extensive current range the voltage drop remains substantially constant During the current increase a glow occurs at the cathode and at the upper end of the range the cathode is completely covered At this point a further current increase can be achieved only if the potential drop across the discharge is increased This portion of the characteristic [illegible] ulative ionization and secondary emission at the cathode See GLOW DISCHARGE

Further increase in current leads to another mode of discharge the arc This is shown as region D in the illustration Characteristic of this mode is the low cathode potential fall and the very high current density It is generally felt that the predominant effect in the production of the large number of electrons at the cathode necessary for the arc is thermionic emission This is consistent with the very high temperatures known to exist either generally or locally on the cathode Although the arc type of discharge has very great commercial value, the mechanism of its operation is not very well understood

In addition to these general types of conduction there are very special cases of considerable interest Some of these are the corona discharge radio-frequency or electrodeless discharge hot cathode discharge and discharges in the presence of a magnetic field [G H M]

Bibliography G P Harnwell *Principles of Electricity and Electromagnetism*, 2d ed, 1949 L B Loeb, *Fundamental Processes of Electrical Discharge in Gases*, 1939, J Millman and S Seely *Electronics*, 2d ed, 1951

Electrical conductivity of metals

Electric currents in metals are caused by mobile relatively free electrons that are not bound to any particular atom and can wander throughout the metal In general, the conductivity of metals is higher than that of other materials and decreases with rising temperature At temperatures near absolute zero, certain metals become superconductors possessing infinite conductivity For an extended discussion of this phenomenon see SUPERCONDUCTIVITY The conductivities of a number of common metals at 0°C are as follows

Metal	*Conductivity,* $(\text{ohm meter})^{-1}$
Silver	66
Copper	64 5
Gold	49
Aluminum	40
Magnesium	25 4
Sodium	23 4
Tungsten	20 4
Potassium	16
Lithium	11 8
Iron	11 5
Cesium	5 2

The current density (current per unit area) in a conductor is proportional to the electric field in the conductor, that is, $\mathbf{J} = \sigma \cdot \mathbf{E}$ where $\mathbf{J}$ is the current density and $\mathbf{E}$ is the electric field intensity The proportionality constant σ is called the electrical conductivity In mks units $\mathbf{J}$ is in amperes per square meter, $\mathbf{E}$ is in volts per meter and σ has the dimensions of $(\text{ohm meter})^{-1}$ In isotropic or nearly isotropic materials such as polycrystalline metals or liquids $\mathbf{J}$ is in the same direction as $\mathbf{E}$, and σ reduces to a scalar constant In this case σ is simply the reciprocal of the resistivity ρ (see RESISTIVITY, ELECTRICAL) In general however, σ must be defined as a tensor, called the conductivity tensor

Electrons possess negative charge hence the direction of current flow is opposite to that of the *flow of electrons* In a solid however electrons may under certain conditions move under the influence of an electric field in such a manner that the net effect is the same as though positively charged carriers of approximately electronic mass were responsible for the current flow It is then common to speak of current due to holes these holes being thought of as charge carriers of positive mass and charge (*see* HOLES IN SOLIDS) Frequently especially in the case of polyvalent metals (and also in semiconductors), the experimental results are described most conveniently by assuming that holes as well as ordinary electrons contribute to charge flow In the theory of conductivity one then speaks of a two band model

In the more general sense the theory of electrical conductivity of metals encompasses all phenomena which relate to the transport of electrons in metals This includes the thermoelectric effects (Peltier Thomson and Seebeck effects) the isothermal magnetic effects (Hall Corbino and magnetoresistance effects) and the thermomagnetic effects (Nernst Ettinghausen and Righi Leduc effects) Moreover since transport of charge is accompanied by transport of electronic mass and energy the general formulation of the theory also contains within its framework the theory of that portion of the thermal conductivity which is due to the presence of mobile electrons or holes in the metal *See* CONDUCTION (HEAT), *see also* BOLTZMANN TRANSPORT EQUATION, FREE ELECTRON THEORY OF METALS HALL EFFECT MAGNETORESISTANCE RELAXATION TIME (ELECTRONS), THERMOELECTRICITY THERMOMAGNETIC EFFECTS WIEDEMANN FRANZ LAW

[F J B]

Electrical degree

A unit equal to 1/360 of a complete cycle of electric current or voltage In an electric machine it is 1/360 of the angle subtended at the axis by two consecutive field poles of like polarity since the voltage wave generated in a conductor completes one cycle when it traverses one pair of poles The term mechanical degree is used to designate the space angle between the two positions about the axis of the machine The number of electrical degrees equals the number of mechanical degrees multiplied by the number of pairs of poles on the machine

[A R E]

Electrical engineering

A branch of engineering dealing primarily with electricity and magnetism and devoted to utilization of the forces of nature and materials for the benefit of mankind (*see* ENGINEERING) Electrical engineering encompasses many phases of other engineering sciences and the physical sciences It includes research invention development design application and education Many phases of electrical engineering are based on applications of higher mathematics

The great advances of the past in electrical engineering are closely associated with certain inventions and discoveries which have made practical uses of electricity and magnetism Throughout the history of electrical engineering there have been eras of accelerated engineering activity that are closely identified with important discoveries by a relatively few scientists and engineers In considering the historical development and present scope of electrical engineering it is convenient to consider five eras of development

First era As early as the latter part of the sixteenth century experimenters were exploring the behavior of static electricity W Gilbert (1540–1603) personal physician to Queen Elizabeth I experimented with electric charges and discharges In 1750 Benjamin Franklin proved that lightning was electrical in nature Neither investigator discovered anything that was significant from the standpoint of the applications of electricity Discovery of the presence of magnetism in certain rocks preceded the earliest knowledge of electricity Such knowledge was common about 600 B C Applications of electrical knowledge were completely absent in this era *See* ELECTRICITY MAGNETISM

Second era The second era had its beginning in electrochemical developments Electrochemical deposition was discovered by W Nicholson and A Carlisle in 1800 and in the same year A Volta discovered the principle of the electric battery The voltaic cell was one of the most important discoveries in the history of the electrical art because it provided a continuous source of appreciable amounts of electric power at reasonably low voltage *See* BATTERY (ELECTRIC) This cell later was used as an essential component of the early communication systems such as the telephone and telegraph

The most significant developments of the second era centered around the field of communications The first United States patent on the electrical telegraph was obtained by J Groat in 1800 The invention of a practical electromagnet was announced by Joseph Henry in 1827 These inventions by Groat and Henry opened the way for a still more significant invention the electromagnetic telegraph The principle of this forerunner of the communications industry was conceived in 1831 proven practical in 1837 and patented in 1840 by Samuel F B Morse

Few developments have had greater impact on American life than Morse's invention His idea paved the way for the first system of electrical communication the telegraph This in turn led to the telephone and later to the wireless telegraph The growth of electrical communications resulted in extensive engineering production of electrical equipment and the birth of an electrical industry adding much to the wealth of the United States and at the same time making possible rapid communications throughout the nation *See* COMMUNICATIO

Third era The discovery of electromagnetic induction by Michael Faraday in 1831 established many principles upon which modern machines function Motors generators transformers and many other electrical devices found in heavy electrical industry were made possible by the discoveries of Faraday The contributions of Faraday in the electrical power industry are comparable to those of Morse in the field of communications

One of the first important developments based on the disclosures of Faraday was the electric dynamo English patent No 1858 describes the principle of operation In the following years many types of dc generators were developed and used commercially The Gramme ring armature was one of the first used in conjunction with a commutator This machine was somewhat inefficient but it provided a source of relatively high voltage at a reasonably large power capacity (up to 100 kw) *See* ELECTRIC ROTATING MACHINERY

With the development of the high resistance carbon filament lamp by Thomas Edison in 1880 the dc generator became one of the essential components of the constant potential lighting system Commercial lighting and residential lighting became practical and the electric light and power industry was born One of the most common uses for *direct current* during this period was for street lighting These lamps were of the carbon arc type Many lamps were operated in series from a constant current generator The generators have long since been replaced by the constant current ac transformer and the lamps have been replaced by low voltage incandescent sodium vapor or fluorescent types but the constant current system of power supply for street lighting still prevails *See* ILLUMINATION

The *first transformer* was announced by L Gaulard and J D Gibbs in 1883 This device probably did more to revolutionize the systems of power transmission than any other The advantages of high voltage low current systems over the low voltage high current systems of power transmission were well known Following the discovery of the transformer power could be generated at low voltages (6 600–13 200 volts) transformed to higher voltages (110 000–287 000 volts) for transmission over great distances (several hundred miles) and then reduced by transformers to lower values for utilization This system of high voltage transmission made possible the generation of electric energy in one part of the country and the utilization of that energy in another part This method of power transmission is of great significance in the development of American industry providing better efficiency and dependability in the utilization of electrical energy It also permits interconnection of power systems *See* ELECTRIC POWER SYSTEMS

The first direct current central station in the United States (Pearl Street Station New York) began operation in 1882 In 1886 the first alternating current station was placed in operation The output of this station was limited essentially to lighting because no suitable alternating current motor was available In 1888 however N Tesla was granted a patent on the polyphase ac induction motor which soon became the most commonly used motor for supplying large amounts of power, in its improved state it is most extensively used today

Early alternating current systems were designed for many different frequencies (25 33⅓, 40 50 60 90 130 and 420 cycles) In 1891 through the efforts of the American Institute of Electrical Engineers studies were made to determine the possibilities of standardizing equipment to standard frequency and voltage ratings with the result that 60 cycles was made the standard frequency in the United States Similar studies in Europe resulted in the selection of 50 cycles This standardization resulted in more universally adaptable equipment and great savings on equipment costs These frequencies are still considered standard and are prevalent today

The power industry made rapid strides in this era but the field of communications was not dormant In 1876 Alexander Graham Bell invented the telephone This device was soon put into use and, as a result another huge industry was established Bell's contribution to the field of science and engineering is one of the greatest in history *See* TELEPHONY

Throughout this period of development few people contributed more than Thomas A Edison His work included research invention development and production His activities extended into chemistry electrical dynamos systems of transmission sound recording and reproduction electrical lamps and many others Perhaps one of his most important discoveries was one that he did not pursue sufficiently to realize its vast importance a discovery known in later years as the Edison effect

Fourth era The fourth era of electrical engineering began with the announcement in 1883 of the Edison effect Edison discovered that when a voltage of proper polarity is applied between two electrodes one hot and one cold that are placed in an evacuated enclosure current flows from the hot to the cold element in an external circuit joining the two This phenomenon was the first indication of thermionic emission of electrons This discovery opened the new field of electronics, which has since grown enormously The discovery resulted from keen observations by Edison while he was pursuing research on incandescent lamps Edison was not searching for this effect, it was a discovery made more or less by accident

Lee DeForest introduced the use of the third element (grid) in the vacuum tube in 1906 This development opened an entirely new field of engineering It made possible new systems of communication and methods of control and indicated the possibility of the multielement tube It provided the basis for future developments in electronics *See* ELECTRONICS

Fifth era The fifth era of electrical engineering can be classified as that of engineering research With production methodology being well established there was rapid expansion in research engineering in the first half of the twentieth century

Industrial research laboratories expanded in size and in number College faculties became increasingly aware of the importance of research to education To administer necessary training in research extensive research laboratories were constructed by American universities Academic appointments have been made of many faculty members who are trained in the systematic pursuit of scientific and engineering knowledge Today research is an essential ingredient in the education of the engineering student the agent by which the student develops his originality his inventive genius and his understanding of the world in which he lives

Research today is a big business no longer carried out by isolated individuals working over long periods It is conducted by highly organized groups of investigators who have been selected because of their competence in certain areas of investigation The lapse of 30 years between invention and production which seemed to prevail in the nineteenth century has been shortened to several years and sometimes to several months a saving in time which can be attributed largely to better systems of communication between scientists and engineers in the engineering profession

The need for better communications between electrical engineers led to the establishment of the American Institute of Electrical Engineers in 1884 This organization had a membership of 52 315 in 1958 Another professional organization the Institute of Radio Engineers was founded in 1913 Its present membership is 70 000 Its purpose is to advance the theory and practice of radio and allied branches of engineering and the related arts and sciences and their application to human need

The references cited in this article will lead the reader to other articles discussing certain branches of electrical engineering in more detail For other areas not previously mentioned *see* CIRCUIT ELECTRIC ELECTRICAL MEASUREMENTS HEATING ELECTRIC

Electrical engineers apply their abilities in other engineering fields that are not strictly electrical There is hardly a field of technology to which electrical engineering has not made a contribution For important related engineering fields *see* COMPUTER CONTROL SYSTEMS INSTRUMENTATION SYSTEMS ENGINEERING [A G CO]

Electrical measurements

The measurement of any one of the many quantities by which the behavior of electricity is characterized The knowledge of the quantitative behavior of electricity is essential to scientific and technical progress Electrical measurements play a major role in industry communications and even in such unrelated fields as medicine

Many electrical measurements can be made with direct indicating instruments merely by connecting the instrument properly in the circuit Thus a voltmeter provides a pointer which moves over a scale calibrated in volts and an ammeter in the same way presents a reading of current in amperes Other direct reading instruments are wattmeters frequency meters power factor or phase angle meters and ohmmeters Many electrical quantities are measured both as instantaneous values and as values integrated over time Some electrical measurements must be made with various specialized devices or systems requiring adjustment or balancing to obtain the measured value Typical of these are potentiometers and bridges in many standard and specialized forms

Because of differences in instruments and techniques it is convenient to divide measurements into direct current (dc) and alternating current (ac) classes

DC measurements In dc circuits the measurement of voltage and current often suffices to define the operation of the circuit The product of the two represents power In the commercial sale of dc electricity the measurement of energy must be made with a dc watt hour meter Occasional use is made of a dc ampere hour meter in battery charging installations

To measure high values of current shunts are used to bypass all but a small fraction of the current around the measuring instrument A newer technique employs a form of saturable reactor energized by ac to measure large direct currents *See* CURRENT MEASUREMENT ELECTRIC ENERGY MEASUREMENT ELECTRIC POWER MEASUREMENT VOLTAGE MEASUREMENT

AC measurements Alternating current circuits involve more variables and hence more measurements than dc circuits The most common measurements are voltage current and power the last requires a wattmeter as ac power cannot always be calculated directly from voltage and current Also measured are frequency and power factor (or phase angle) and sometimes waveform or harmonic content Energy is measured by means of the ac induction watt hour meter In general ac instruments differ in principle and design from dc instruments although many ac instruments may be used to measure dc quantities Direct current instruments do not respond to ac quantities but some may be adapted by the addition of rectifiers to convert the ac to dc The thermocouple is another form of convertor by which a dc instrument may be made to read ac quantities *See* PHASE ANGLE MEASUREMENTS

If alternating voltages and currents above the normal ranges of self contained instruments are to be measured instrument transformers may be used to extend the ranges of those instruments In the study of ac waveform a qualitative evaluation may be made with an oscillograph or a cathode ray oscilloscope Quantitative measurement of harmonic content requires the use of a harmonic analyzer *See* HARMONIC ANALYZER INSTRUMENT TRANSFORMER

Accuracy of measurements Accuracy denotes the degree of compliance of the instrument reading with the true value of the measured quantity It i common to describe the instrument s accuracy
 the maximum allowable error Th

instrument with a maximum error of 2% is often described as having an accuracy of 2%. For many applications a small panel or miniature electric instrument with a maximum error of 2.5% of full scale calibration will suffice. More refined instruments are available with maximum errors of 1%, ½% or ¼%. When measurements of higher accuracy than this are required, measurement systems such as potentiometers and bridges must be used. Direct current and voltage can readily be measured in this way to an accuracy of 0.01%. Alternating current measurements can be readily made to an accuracy of 0.1% or better.

Laboratory measurements. In a laboratory emphasis is normally placed on accuracy and on completeness of facilities to deal with all types of measurements. There is relatively little limitation on the size and complexity of equipment used. If standardizing service is a function of the laboratory, the equipment must include standard cells and precision standard resistors and also suitable potentiometers, bridges, shunts, and volt boxes (voltage-dividing resistors to extend the range of voltage measurements). With these the calibration of dc instruments can be performed with high accuracy. Extension of calibration service to ac instruments requires transfer standards (instruments having negligible difference in performance when oper and

in generating stations and substations, service shops, factory testing areas, ships and aircraft. For these uses equipment is chosen to perform only specialized services. Accuracy well below that of laboratory measurements is usually permissible. Convenience, compactness, and often portability are prime considerations in choosing equipment. Electric instruments for this kind of service are often of the panel type, sometimes in miniature sizes. Multipurpose and multirange instruments like the

ating current and voltage at power frequencies, the hook-on voltammeter provides readings within 2% maximum error. As a *voltmeter* it is connected to the circuit with spring clips, as an ammeter it operates on the current transformer *principle* (*see* INSTRUMENT TRANSFORMER). The core, which is circular and

the circuit under measurement, the conductor itself *becomes* a one-turn primary winding in the transformer measurement system. The measurement of power can also be made with a hook-on wattmeter.

On power systems field measurements may be desired continuously over a *period* of time, sometimes at unattended locations. For this purpose recording instruments may be used or the reading may be telemetered to a manned *station* (*see* RECORDING INSTRUMENTS, GRAPHIC; TELEMETERING). Any instrument that is made in indicating form can be made in recording form, but the greater power necessary to drive a marking device over a chart may call for some kind of amplification.

The operation of a power system also calls for the recording of disturbances due to lightning strokes, insulation flashover, short circuits, and other transient phenomena. Recording oscillographs used for this purpose can be triggered by any condition that deviates from normal operation, thus making a record of the disturbance.

Frequency considerations. The measurement of voltage and current is commonly made over a frequency range from a few cycles per second (cps) up through 2000 megacycles (Mc). The frequency at which measurements are to be made dictates the type of equipment needed, the precautions to be taken, and the degree of accuracy which may reasonably be expected. Alternating-current instruments of the moving-iron, fixed-coil type are intended primarily for 60-cycle applications but may be used with only moderate errors up to several hundred cycles. Electrodynamic (moving-coil, fixed-coil) instruments, which are generally of greater accuracy, may also be used in this frequency range. Errors in such instruments result mainly from reactance effects, which may be minimized by special design to permit operation to several kilocycles (kc). Rectifier-type instruments possess only small frequency errors up to several kilocycles in relation to their over-all accuracy rating, which is of the order of 2.5% error. Vacuum-type voltmeters, which are of the same general accuracy, are especially suited for use over a wide range of frequencies.

Circuit loading. All electric instruments draw some power from the circuits to which they are connected, and ac instruments generally take more power than dc instruments. This circuit loading may alter appreciably the quantity being measured. For instance, an ac voltmeter rated 150 volts and having a resistance of 3000 ohms is perfectly suitable to measure voltage on a 120-volt house lighting circuit. However, if the same voltmeter is connected to the terminals of a small power amplifier with a maximum output of 200 milliamperes, the voltmeter will load the source and seriously reduce its voltage. To avoid this error the measurement should be made with a rectifier voltmeter (about 150,000 ohms resistance) or a vacuum-tube voltmeter (above 0.5 megohms resistance).

In the measurement of current, consideration must be given to the voltage drop in the ammeter. If it is appreciable in relation to the source voltage, the current is not the same as that if the ammeter were not connected in the circuit. The magnitude of this error can usually be evaluated and *minimized* by proper choice of instruments. [I.F.K.]

Time dependence. When voltage and current are variable functions of time, the measurement of frequency, wavelength, and waveform is of importance in addition to phase angle measurements

At frequencies where the physical dimensions of the electric circuit are small compared to the wave length of the voltage and current the frequency is said to be low and various forms of frequency meters are employed depending upon the range involved

At higher frequencies it often becomes necessary to measure frequency and wavelength independently since their product is not always a constant under this condition *See* FREQUENCY MEASUREMENT, WAVELENGTH MEASUREMENT

The waveform of the electrical quantity being measured is of importance Many indicating instruments are calibrated to give correct readings only for sine-wave inputs If the waveform is nonsinusoidal it is necessary to consider the principles of operation of the particular instrument to interpret the meter indication correctly For example an electronic voltmeter which measures peak values but is calibrated in root mean square (rms) values based on sinusoidal wave shape would not give correct rms values for a nonsinusoidal wave *See* WAVEFORM DETERMINATION

Measurement of parameters The parameters of any electric circuit are the resistances inductances and capacitances along and between the conducting branches of the circuit including any ground plane that may be near or surrounding the circuit The measurement of these parameters may be classified according to the apparent disposition of the parameters which is a function of frequency For any given circuit there is some frequency below which the circuit can be treated as having lumped parameters or circuit elements above this frequency the parameters must be considered as being distributed throughout the circuit

Lumped parameters The measurement of lumped parameters may be subdivided according to the measurement accuracy desired If errors of several per cent are permissible direct reading instrument which indicate the value of the parameter directly on a calibrated meter scale are available Inductance and capacitance measurements are made at some convenient frequency Of this class of instruments the ohmmeter used for measuring resistance at zero frequency (direct current) is the only one in common use

For greater accuracies bridge measurements are preferred Direct current measurements are made with the Wheatstone bridge with maximum errors on the order of 0 01% Resistances of less than a few ohms can be satisfactorily measured only with a bridge regardless of the accuracy desired See RESISTANCE MEASUREMENT

Most circuit designers prefer to measure inductance and capacitance in the particular operating frequency range under consideration and the ac bridge is commonly used Bridge measurements provide numerous advantages over other methods including high accuracy and the ability to compare the unknown to a known standard Bridges are designed to operate in various frequency ranges from direct current to several hundred kilocycles from several kilocycles to several megacycles from 1 or 2 to several hundred megacycles, and from several hundred to several thousand megacycles At the higher frequencies the application of the bridge becomes more complicated and considerable caution and planning is necessary if reliable results are to be obtained *See* CAPACITANCE MEASUREMENT IMPEDANCE MEASUREMENTS HIGH FREQUENCY INDUCTANCE MEASUREMENT

A unique instrument known *as a Q meter* is available for measuring inductance or capacitance and effective resistance at *radio frequencies See* Q METER [R L R]

Distributed parameters Electrical systems can be completely described by their associated electric and magnetic fields the properties of the materials involved the physical dimensions, and the velocity of light When the dimensions are small compared with the wavelength however it is more convenient to treat them as circuits composed of lumped parameters

At low frequencies lumped inductance and capacitance can be used although they are rigorously derived only for nonvarying currents and voltages, respectively At high frequencies the finite propagation velocity of electromagnetic waves cannot be [illegible]

ditions permit a rigorous definition of distributed inductance and capacitance These distributed parameters combine with the resistance of a pair of conductors and the conductance between them to define the behavior of a transmission line for plane wave propagation and to relate the voltage between conductors at any point on the line to the voltage at any other point *See* TRANSMISSION LINES

The concept of distributed parameters is also useful at low frequencies when it must be recognized that a circuit component nominally representable by a single parameter is actually modified by the presence of residual parameters Thus a coil has not only inductance but capacitance and resistance as well This capacitance is definable by low frequency analysis since the dimensions are small but it cannot be localized and represented as a unique lumped parameter because the winding is not an equipotential surface

For example a coil mounted over a ground plane has one terminal grounded The voltage between winding and ground increases from zero at the grounded terminal to maximum at the other Capacitance near the grounded terminal is therefore less effective than capacitance near the other The resultant effective terminal capacitance is in consequence only one third of the total capacitance for uniformly distributed capacitance For other conditions of grounding the ratio of effective capacitance to total distributed capacitance will again be different

Values of distributed parameters are inferred from the behavior of the system that they define

For transmission lines measurements may involve observing the voltage distribution along the line under different terminal conditions For circuit elements impedance may be measured at different frequencies or under different conditions of adjustment of some known lumped parameter For the various methods of measurement *see* IMPEDANCE MEASUREMENTS HIGH FREQUENCY [D B S]

Bibliography R F Field and D B Sinclair A method for determining the residual inductance and resistance of a variable air condenser at radio frequencies *Proc IRE* 24(2) 1936 E I Ginzton *Microwave Measurements* 1957 F K Harris *Electrical Measurements* 1952 F A Laws *Electrical Measurements* 1938 G R Partridge *Principles of Electronic Instruments* 1958 F E Terman and J M Pettit *Electronic Measurements* 1952

Electrical standards

The standards in terms of which electrical quantities are evaluated The ohm and the volt are now the fundamental electrical standards

In the United States the establishment and maintenance of all scientific standards is the responsibility of the National Bureau of Standards The legal standards have been established by acts of Congress The ultimate calibration of all instruments must be against the standards maintained at the Bureau of Standards but for convenience laboratories and instrument manufacturers usually possess their own secondary standards The calibration of instruments is usually accurate enough so that recalibration by the user is necessary only in the most precise applications

Systems Historically two sets of electrical standards have been used The older known as the international system was based upon the resistance of a mercury column of prescribed dimensions and upon a current which would deposit silver at a prescribed rate from an electrolyte It was hoped that these standards could be accurately reproduced at any time and place by following simple experimental procedures Actually it was soon found that the accuracy with which the standards could be experimentally duplicated was insufficient for the needs of science and technology Furthermore the difference between the electrical watt defined by this system and the mechanical watt was undesirably large The so called international system of units was finally abandoned by world wide agreement on January 1 1948 in favor of a so called absolute system based directly upon the mechanical units of mass length and time

Absolute standards In the absolute system of electrical standards the kilogram meter and second are taken as the mechanical standards In order to include electrical quantities a fourth fundamental standard must be chosen This may conveniently be taken to be the permeability of free space μ This is assigned the value of $4\pi \times 10^{-7}$ mks units (newtons per ampere squared) Primary standards of resistance (the absolute ohm) and of current (the absolute ampere) are established from combined electrical and mechanical measurements based upon these four fundamental standards Subsequently all other electrical standards can be defined in terms of the ohm and the ampere

The resistance of a standard resistor in absolute ohms can be determined by comparison with a standard inductor by means of an appropriate ac bridge circuit The inductance L of the inductor can be calculated from the geometry of the coil and the permeability μ of the surrounding medium The permeability of air is very nearly μ_0 the small correction can be determined independently of any particular system of units The inductive reactance X_L presented by the coil to alternating current of frequency f cycles per second is given by $X_L = 2\pi fL$ absolute ohms Thus when the ac bridge is balanced the resistance R of the resistor being calibrated is given by the relation $R = 2\pi fL$

To establish the absolute ampere the force between two parallel conductors is measured by a current balance The magnitude of the force depends upon the current in the conductors and upon the permeability of the space between them The force between two wires one meter apart each carrying a current of one absolute ampere is taken to be 2×10^{-7} newton per meter of length

Once the absolute ohm and the absolute ampere have been established the standard of potential is determined by the IR drop across a standard resistor The electromotive force (emf) of a standard cell in absolute volts can be measured by means of a potentiometer The standards which are more readily preserved are the absolute ohm (standard resistor) and the absolute volt (standard cell) rather than the absolute ampere

While the sizes of the absolute units of resistance and current depend upon the particular unit system and thus upon the value assigned to μ_0 the product I^2R (power) is independent of μ_0 The absolute electrical watt is equal to the mechanical joule per second

International standards The establishment of the International standards in 1894 is now mainly of historical interest In this system the ohm and the ampere were the fundamental standards and depended upon the properties of particular substances So long as the required precision is not too high, these standards are more easily reproducible than the absolute standards However accurate measurements have shown discrepancies of the order of 0.05% between the two sets of units as follows

United States standards. The actual calibrations of a standard resistor in absolute ohms and of a standard cell in absolute volts require a great deal of experimental skill and patience It is economically feasible to carry out such measurements only at intervals of several years During the intervening periods the standards are preserved as special groups of wire wound resistors and standard cells maintained at the Bureau of Standards These constitute the primary electrical standards of the United States All values for other electrical quantities are determined from these In addition a number of secondary standards are permanently maintained

Standards of resistance The primary standard is a group of ten 1-ohm wire wound Thomas type resistors built in 1933 They consist of No 12 American Wire Gage manganin wire, vacuum annealed at 550°C, and for protection sealed in air in double-walled containers The resistors are compared one against another each year Through 1958 the maximum relative change in any one of the group of ten referred to the average of this group was about 2×10^{-6} The average change of individuals with respect to the group mean was less than 1×10^{-6}

Manganin (composed roughly of 80% copper 15% manganese and 5% nickel) is eminently suitable for standard resistors because its temperature coefficient of resistivity is only a few parts in 10^6 and its thermal emf against copper is only 2 3 microvolts per degree centigrade (μv/°C) This means that resistance comparisons to seven significant figures are feasible Secondary standards of resistance in the range 10 microhms to 1 megohm can be set up in terms of parallel or series combinations of the primary standards

Manganin has the disadvantage that its resistivity is sensitive to internal strains, for this reason the wire must be well annealed before use Furthermore it oxidizes to some extent It is essential to avoid strains while the resistors are in use and also to exclude moisture The special mountings of the Bureau of Standards resistors protect against both of these effects The secondary standards of large resistance are more susceptible to slow changes due to oxidation because of their larger surface to volume ratio

Standards of emf The primary standard consists of a group of 47 saturated cadmium (Weston) cells maintained at a temperature of 28 00 ± 0 01°C Of these cells 33 are of the acid type (0 03–0 05 *N* sulfuric acid), and the remaining 14 are neutral The emf of the acid cells is 20–30 μv lower than that of the neutral cells This difference must be allowed for in the calibration of secondary standards The cells are frequently intercompared From time to time new ones are prepared and calibrated so that when one of the primary cells wears out it can be replaced by a cell with an emf whose drift with time is already known The standard of potential is taken from the average of all 47 cells An international comparison made in 1957 indicated that at that time the United States standard differed by 1 3 μv from that of the International Bureau of Weights and Measures at Sevres, France Over a period of time the reproducibility of the primary standard is of the order of 1 μv

Secondary standards of emf are usually Weston cadmium cells of either the saturated or unsaturated type The former are more stable over long periods of time, but the unsaturated cells are more portable and are less sensitive to changes in temperature Many laboratory standards are of the unsaturated acid type For most work the emf of these may be taken as 1 0183 volts at 20°C

An unknown voltage must be determined by comparison with the emf of a standard cell by some sort of null instrument such as a potentiometer No more than 100 microamperes should ever be drawn from a standard cell Even momentary short circuiting may cause permanent damage Furthermore all standard cells should be kept away from bright light and protected from sudden or large changes in temperature *See* ELECTROMOTIVE FORCE (CELLS), WET CELL

Every research laboratory should possess some high quality standard resistors and several Weston cells Unless the work is of a very special nature, these will be adequate standards against which to calibrate other electrical measurements

Standards of capacitance and inductance The Bureau of Standards also maintains secondary standards of capacitance and inductance Absolute values are calculated from the geometry, and from μ_0 and ϵ_0, permittivity of empty space Other elements can be calibrated against these standards by means of ac bridge circuits

Very accurate capacitance standards covering the range from 10^{-6} $\mu\mu$f to 1 μf are now available Unknown capacitances can be calibrated to within a few parts in 10^6 by comparison with these new standards A special type of bridge circuit is used Accurate standards with capacitance greater than 1 μf cannot be made because of high dielectric losses in the materials which must be used Air gap capacitors in this range are prohibitively large

Standard inductors of high accuracy have been built at the various standards laboratories for absolute ohm determinations These same coils can be used in the calibration of unknown inductances

Standard inductors can be made with or without iron cores Losses are kept to a minimum by suitable laminations In any event, corrections have to be made for eddy current losses

Frequency standards The frequency of commercial 60 cycle current is reliable only to about 1%, [illegible] extended period of time the [illegible]

[illegible]

Electrically driven tuning forks are reliable to 0 1% and are a commonly used laboratory standard Precision forks with temperature control give frequencies accurate to 1 part in 10^5 or 1 part in 10^6 if correction is made for changes in pressure Tuning forks can give fund

quencies up to several times 10^4 cps and higher frequencies are readily calibrated against the harmonics of the standard frequencies

The primary frequency standard in the United States is a 100 kilocycle quartz crystal oscillator maintained under conditions of constant pressure and temperature by the National Bureau of Standards This is frequently checked against Naval Observatory time Broadcasts of audio signals of frequency known to 1 part in 50×10 and time signals accurate to 1 μsec are made over the Bureau of Standards short wave radio stations WWV and WWVH

Alternative frequency standards are furnished by certain vibrating atomic systems in particular the ammonia molecule and the cesium atom The NH_3 frequency 23 870 megacycles (Mc) is known to 1 part in 10 or better while the cesium frequency 9192 Mc is certain to one part in about 10^{9} (*see* ATOMIC CLOCK) Such systems should eventually replace the rate of rotation of the earth as the basic time and frequency standards The length of the day is subject to irregular variations of the order of 1 part in 2×10^7 and is secularly increasing at the rate of 10^{-3} sec per century *See* TIME *see also* ELECTRICAL UNITS [J W ST]

Bibliography F K Harris *Electrical Measurements* 1952 F K Harris Electrical standards in D E Gray (ed) *American Institute of Physics Handbook* 1957

Electrical units

The quantities adopted as standards of measurement in electricity for example the ohm is a unit of electrical resistance Magnetic units are also treated in this article

Any system of units for electricity and magnetism must necessarily include mechanics as well since mechanical quantities play important roles in nearly all electromagnetic phenomena At the present time there are four systems of electrical and magnetic units in common use In addition several systems which have never won wide acceptance have been proposed for special applications

The problem of establishing a system of electrical units is analogous to the more familiar situation in mechanics In the latter field in the so called absolute systems units of mass length and time are chosen as fundamental These choices are made because of the convenience of establishing and maintaining standards The unit of force is then defined from Newton s second law force = mass × acceleration On the other hand in the engineering systems force length and time are the usual choices of fundamental units Mass is treated as a secondary quantity and is measured in the same units as force Then Newton s second law must be written $F = ma/g$ where g is the acceleration of gravity This is an illustration of the frequently occurring dependence of mathematical expressions of laws of physics upon the choice of units

In setting up any system of units it is necessary to make certain choices of fundamental quantities These choices are completely arbitrary and are made purely on the basis of convenience *See* DIMENSIONAL ANALYSIS, UNITS SYSTEMS OF

In order to include electrical as well as mechanical quantities at least one additional choice of a fundamental unit must be made This selection involves an arbitrary assumption about the magnitude and the units of some one electrical quantity

All four of the systems of electrical units in common use are based upon either the meter kilogram second (mks) system or the centimeter gram second (cgs) system of mechanical units Because of its convenience for practical calculations the rationalized mks system of electrical units is now the most widely used

The additional arbitrary choices for each of the four systems are explained below

The rationalized mks system The basic unit is the ampere which is defined in terms of the force of attraction between two neighboring currents. Two parallel conductors 1 meter (m) apart each carrying a current of 1 ampere (amp) are postulated to exert on each other a force of 2×10^{-7} newton (per meter of length) This value was chosen so that the ampere as defined in this manner would have the same magnitude as with the earlier definition in terms of the electrolytic deposition of silver Actually there is a small discrepancy (*see* ELECTRICAL STANDARDS) The newer definition is equivalent to assigning the value $4\pi \times 10^{-7}$ newton/amp^2 to the permeability of empty space (μ_0)

The cgs electrostatic system The unit of charge is defined as being such that two unit positive charges 1 centimeter (cm) apart (in vacuum) repel one another with a force of 1 dyne This is equivalent to assuming the permittivity of empty space (ϵ_0) to be a pure number of magnitude unity This is the most convenient unit system for treating purely electrostatic problems

The cgs electromagnetic system The unit of magnetic pole strength is defined as being such that two unit north poles 1 cm apart (in vacuum) repel one another with a force of 1 dyne This is equivalent to assuming $\mu_0 = 1$ (a pure number) for empty space Since it is now believed that all magnetic effects are produced by moving charges and that free magnetic poles do not exist this system is much less commonly used now than formerly The ratio of the electromagnetic unit of charge to the electrostatic unit of charge is equal to the velocity of light c (3×10^{10} cm/sec)

The Gaussian system This system is a combination of the two cgs systems just described Electrostatic definitions are used for all electrical quantities and electromagnetic definitions for all magnetic quantities Both ϵ_0 and μ_0 are unity and dimensionless These choices introduce factors of c into many of the relations The Gaussian system is the most convenient of the four for applications in some branches of theoretical physics

Electrical and magnetic units

Quantity	Rationalized mks system unit	cgs electrostatic system: Unit	cgs electrostatic system: Equivalent in mks	cgs electromagnetic system: Unit	cgs electromagnetic system: Equivalent in mks	Gaussian system: Unit	Gaussian system: Equivalent in mks
Mass	kilogram	gram	10^{-3}	gram	10^{-3}	gram	10^{-3}
Length	meter	centimeter	10^{-2}	centimeter	10^{-2}	centimeter	10^{-2}
Time	second	second	1	second	1	second	1
Charge	coulomb	statcoulomb	3.33×10^{-10}	abcoulomb	10	statcoulomb	3.33×10^{-10}
Current	ampere	statampere	3.33×10^{-10}	abampere	10	statampere	3.33×10^{-10}
Potential	volt	statvolt	3×10^{2}	abvolt	10^{-8}	statvolt	3×10^{2}
Resistance	ohm	statohm	9×10	abohm	10^{-9}	statohm	9×10
Resistivity	ohm meter	statohm-cm	9×10^{9}	abohm-cm	10	statohm-cm	9×10^{9}
Power	watt	erg/sec	10	erg/sec	10	erg/sec	10^{-7}
Electric field	volt/meter or newton/coulomb	statvolt/cm or dyne/statcoulomb	3×10^{4}	abvolt/cm or dyne/abcoulomb	10^{-6}	statvolt/cm or dyne/statcoulomb	3×10
Electric displacement	coulomb/meter²	statcoulomb/cm²	2.65×10	Seldom used	7.96×10^{3}	statcoulomb/cm²	2.66×10
Capacitance	farad	statfarad	1.11×10^{-1}	abfarad	10^{9}	statfarad	1.11×10^{-2}
Inductance	henry	stathenry	9×10	abhenry	10^{-9}	abhenry	10^{-9}
Dipole moment	coulomb-meter	statcoulomb-cm	3.33×10^{-1}	abcoulomb-cm	10	statcoulomb-cm	3.33×10^{-1}
Polarization	coulomb/meter²	statcoulomb/cm²	3.33×10^{-1}	abcoulomb/cm²	10	statcoulomb/cm²	3.33×10^{-6}
Magnetic pole	weber	Seldom used	3.77×10^{3}	unit pole	1.257×10	unit pole	1.257×10
Magnetic moment	weber meter	Seldom used	37.7	pole-cm	1.257×10^{-9}	pole-cm	1.257×10^{-9}
Magnetization	weber/meter²	Seldom used	3.77×10	pole/cm²	1.257×10^{-3}	pole/cm²	1.257×10^{-3}
Magnetic field	amp-turn/m	Seldom used	2.65×10^{-9}	oersted	79.6	oersted	79.6
Magnetic induction	weber/meter²	Seldom used	3×10^{6}	gauss	10	gauss	10^{-4}
Magnetic flux	weber	Seldom used	3×10^{2}	maxwell	10^{-8}	maxwell	10^{-8}
Magnetomotive force	ampere-turn	Seldom used	2.65×10	gilbert	0.796	gilbert	0.796
Reluctance	amp-turn/weber	Seldom used	8.84×10	gilbert/maxwell	7.96×10	gilbert/maxwell	7.96×10
Permittivity	coulomb/newton-m²	Dimensionless (=1 in vacuum)	*	sec²/cm²	*	Dimensionless (=1 in vacuum)	*
Permeability	newtons/amp²	sec²/cm²	*	Dimensionless (=1 in vacuum)	*	Dimensionless (=1 in vacuum)	*
Dielectric constant	Dimensionless	Dimensionless	1	Dimensionless	1	Dimensionless	1
Relative permeability	Dimensionless	Dimensionless	1	Dimensionless	1	Dimensionless	1

* In the rationalized mks system the permittivity of empty space ϵ_0 is defined as 8.84×10^{-12} coulomb/newton m², the permeability of empty space μ_0 is 1.257×10^{-6} newton/amp² (henries/m). In any system $\sqrt{\mu_0\epsilon_0} = 1/c^2$

In the table are listed the principal electrical and magnetic units in each of the four systems. The numerical values given are the equivalents of each unit in terms of the corresponding unit in the rationalized mks system.

As an illustration of the use of the table consider units of potential. In the rationalized mks system the unit is the volt. In the cgs electrostatic and Gaussian systems the unit is the statvolt (1 statvolt = 300 volts). In the cgs electromagnetic system the unit is the abvolt (1 abvolt = 10^{-8} volt). The conversion factors in the table are not in general pure numbers. The dimensional formulas of the various quantities depend upon the particular choice of unit system.

The mks system is said to be rationalized because of the factor 4π which is included in Coulomb's law

$$F = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2}$$

F is force in newtons between charges q_1 and q_2 coulombs separated by a distance of r meters. If the factor of 4π were omitted here it would appear in formulas which are used more often. Unrationalized mks units are employed only rarely.

The electrical units just discussed are inconveniently large for atomic physics. The charge on the electron is only 1.6×10^{-19} coulomb or 4.8×10^{-10} statcoulomb. Atomic diameters are of the order of 10^{-10} m. In order to have more convenient magnitudes for the various quantities, various systems of atomic units have been proposed. For example one possibility which has been suggested is to take the unit of charge to be the electronic charge, the unit of mass to be the mass of the electron, the unit of length to be the radius of the first Bohr orbit of hydrogen, and the unit of time to be the classical period of rotation of the electron in the first Bohr orbit. This is only one of a number of such systems. None has won wide acceptance. For detailed discussion of the more common electrical units see articles under the individual names. [J W ST]

Electricity

Electricity comprises those physical phenomena involving electric charges and their effects when at rest and when in motion. Electricity is manifested as a force of attraction independent of gravitational and short-range nuclear attraction when two oppositely charged bodies are brought close to one another. It is now known that the elementary (nondivisible) electric charges are possessed by electrons and protons. The charge of the electron is equal in magnitude to that of the proton but is electrically opposite. The electron's

charge is arbitrarily termed negative and that of the proton positive Magnetism those physical phenomena involving magnetic fields and their effects upon materials manifests itself in the presence of moving electric charge For this reason magnetism was originally considered to be a part of electricity See CHARGE ELECTRIC MAGNETISM

Historical development The earliest observations of electric effects were made on naturally occurring substances Magnetism was observed in the attraction of metallic iron by the iron ore magnetite The natural resin amber was found to become electrified when rubbed (triboelectrification) and to attract lightweight objects Both of these phenomena were known to Thales of Miletus (640–546 B C) Jerome Cardan in 1551 first clearly distinguished the difference between the attractive properties of amber and magnetite thus presaging the division of electric and magnetic effects He al o envisioned electricity as a type of fluid a viewpoint that was developed more extensively in the late eighteenth and early nineteenth centuries In 1600 W Gilbert observed variations in the amounts of electrification of various substances He divided substances into two classes according to whether they did or did not electrify by rubbing. The division actually is into good and poor conductors respectively A two fluid theory was first proposed by C F duFay in 1733 A one fluid theory of electricity was propounded in 1747 by Benjamin Franklin who called an excess of the fluid positive electrification and a deficiency of fluid negative electrification This theory fell into disrepute but the choice of positive and negative remains Although fluid theories of electricity were superseded at the end of the nineteenth century the concept of electricity as a substance persists

The quantitative development of electricity began late in the eighteenth century J B Priestley in 1767 and C A Coulomb in 1785 discovered independently the inverse square law for stationary charges This law serves as a foundation for electrostatics See COULOMB S LAW ELECTROSTATICS

In 1800 A Volta constructed and experimented with the voltaic pile the predecessor of modern batteries It provided the first continuous sources of electricity In 1820 H C Oersted demonstrated magnetic effects arising from electric currents The production of induced electric currents by changing magnetic fields was demonstrated by M Faraday in 1831 Faraday in 1851 also proposed giving physical reality to the concept of lines of force This was the first step in the direction of shifting the emphasis away from the charges and onto the associated fields See ELECTROMAGNETISM INDUCTION ELECTROMAGNETIC

In 1865 J C Maxwell presented his mathematical theory of the electromagnetic field This theory proposed a continuous electric fluid It remains valid today in the large realm of electromagnetic phenomena where atomic effects can be neglected Its most radical prediction the propagation of electromagnetic radiation was convincingly demonstrated by H Hertz in 1887 Thus Maxwell's theory not only synthesized a unified theory of electricity and magnetism but also showed optics to be a branch of electromagnetism See ELECTROMAGNETIC RADIATION, MAXWELL'S EQUATIONS

The developments of theories about electricity subsequent to Maxwell have all been concerned with the microscopic realm Faraday's experiments on electrolysis in 1833 had indicated a natural unit of electric charge thus pointing towards a discrete rather than continuous charge Thus the groundwork for exceptions to Maxwell s theory of electromagnetism was laid even before the theory was developed H A Lorentz began the attempt to reconcile these viewpoints with his electron theory in 1895 He postulated discrete charges called electrons The interactions between the electrons were to be determined by the fields as given by Maxwell s equations The existence of electrons negatively charged particles was demonstrated by J J Thomson in 1897 using a Crookes tube The existence of positively charged particles (protons) was shown shortly afterwards (1898) by W Wien who observed the deflection of canal rays Since that time 11 other particles having charges numerically equal to that of the electron have been found Of these 13 particles only two the electron and the proton exist in a stable condition on earth See ELECTRON ELEMENTARY PARTICLE, PROTON

A second departure from classical Maxwell theory was brought on by M Planck s studies of the electromagnetic radiation emitted by black bodies These studies led Planck to postulate that electromagnetic radiation was emitted in discrete amounts called quanta This quantum hypothesis ultimately led to the formulation of modern quantum mechanics The most satisfactory fusion of electromagnetic theory and quantum mechanics was achieved in 1948 with the work of J Schwinger and R Feynman in quantum electrodynamics These formulations suppress the particle aspect and emphasize the field throughout See HEAT RADIATION, QUANTUM ELECTRODYNAMICS QUANTUM MECHANICS

Sources of electricity The sources of electricity in modern technology depend strongly on the application for which they are intended

The principal use of static electricity today is in the production of high electric fields Such fields are used in industry for testing the ability of components such as insulators and condensers to withstand high voltages and as accelerating fields for charged particle accelerators The principal source of such fields today is the Van de Graaff generator See PARTICLE ACCELERATOR VAN DE GRAAFF GENERATOR

The major use of electricity today arises in devices using electric currents alternating at low or zero frequency The use of alternating current introduced by S Z de Ferranti in 1885–1890 allows power transmission over long distances at very high voltages with a resulting low percentage power loss followed by highly efficient conversion

to lower voltages for the consumer through the use of transformers Large amounts of zero-frequency current, that is, direct current, are used in the electrodeposition of metals both in plating and in metal production, for example in the reduction of aluminum ore To avoid power transmission difficulties such facilities are frequently located near sources of abundant power *See* ALTERNATING CURRENT, CURRENT ELECTRIC, DIRECT CURRENT

The principal sources of low frequency electricity are rotary generators whose operation is based on the Faraday induction principle The force to drive such generators derives from the flow of water or the expansion of gases as in steam and internal combustion engines The heat which causes the gas expansion derives principally from fossil fuels Fission reactors are being used increasingly as heat sources, although they are not economically competitive with fossil fuels in all areas To a limited extent, steam from such natural sources as geysers is being used in Italy and has been projected in the United States The development of fusion reactors that could act as heat sources is being attempted

A more direct method of using fission or fusion reactors is the direct conversion of the energy released in the nuclear process into electricity This has been achieved on a laboratory scale in the case of f [illegible]

[illegible] ca ions equipment, television, and radar, involve the consumption of only moderate amounts of power generally derived from low frequency sources If the power requirements are moderate and portability is needed, the use of ordinary chemical batteries is possible Ion permeable membrane batteries are a later development in this line The successful use of thermoelectric generators based on the Seebeck effect in semiconductors has been reported in Russia and in the United States In a particularly compact low power device constructed in the United States the heat needed for the operation of such a generator has been supplied by the energy release in the radioactive decay of suitably encapsulated isotopes produced in fission reactors *See* BATTERY (ELECTRIC), ION PERMEABLE MEMBRANE, THERMOELECTRICITY

The Bell solar battery, also a semiconductor device has been used to provide charging current for storage batteries in telephone service and in communications equipment in artificial satellites *See* SOLAR BATTERY

There are a number of other effects which might also serve to convert various forms of energy into electrical energy, but they do not seem generally practicable

The changing magnetic flux required for the Faraday induction may be produced by an oscillating (rather than rotary) mechanism or by varying the temperature of a magnetic circuit whose components are made of a substance with a highly temperature dependent permeability It has been proposed to extract the energy of the fission (or possibly the fusion) reaction directly by inducing currents in external circuits by the changing magnetic field of bursts of ions from the reaction

Direct conversion of mechanical energy into electrical energy is possible by utilizing the phenomena of piezoelectricity and magnetostriction These have some present application in acoustics and stress measurements Pyroelectricity is a thermodynamic corollary of piezoelectricity *See* MAGNETOSTRICTION PIEZOELECTRICITY, PYROELECTRICITY

Another set of sources of electricity are those in which charged particles are released with some energy and collected in some manner Charged particles are suitably released in radioactive decay in the photoelectric effect and in thermionic emission among other ways The photovoltaic effect may also be in this set *See* ELECTRON EMISSION, PHOTOELECTRICITY

The differences of work functions of various materials can be used for energy conversion The contact potential difference may be used to convert heat directly to electricity or to provide improved collection for currents arising from some other source such as radioactivity *See* CONTACT POTENTIAL DIFFERENCE WORK FUNCTION (ELECTRONIC)

Other possible sources of electricity arise from the existence of electrokinetic potentials in flowing fluids and of phase transition potentials such as occur in the Workman Reynolds effect The possibilities of combining several effects also exist as exemplified in thermogalvanic potentials It also appears that organic materials (as distinguished from the inorganic materials for which most of the work already described was done) merit investigation A primitive type of organic solar battery has been developed *See* ATMOSPHERIC ELECTRICITY, BIOPOTENTIALS AND ELECTROPHYSIOLOGY, CIRCUIT ELECTRIC, CONDUCTION (ELECTRICITY), ELECTRIC POWER MEASUREMENT ELECTRICAL ENGINEERING, ELECTRICAL UNITS, ELECTROCHEMISTRY, ELECTRONICS, TERRESTRIAL ELECTRICITY [W A R]

Bibliography B I Bleaney and B Bleaney *Electricity and Magnetism*, 1957 G P Harnwell *Principles of Electricity and Electromagnetism*, 2d ed , 1949 L B Loeb *Fundamentals of Electricity and Magnetism*, 3d ed , 1947

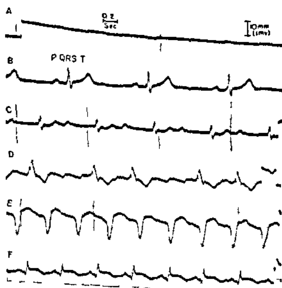
Electroacoustics

That part of electronics concerned with the conversion of acoustical energy into electrical energy or vice versa The means for conversion from acoustical to electrical systems and vice versa are termed electroacoustic transducers or electroacoustic systems Electroacoustic systems are used in the telephone, phonograph radio sound motion picture and magnetic tape reproducers and sound reinforcing systems for converting acoustical waves at the input into corresponding electrical waves and for converting electrical waves at the output into corresponding sound waves *See* SOUND REPRODUCTION SYSTEMS, ELECTRICAL [H F O]

Electrocardiography

The science of recording and interpreting the electrical manifestations of the heart beat The electrocardiogram (ECG or EKG) is the record obtained the electrocardiograph the recording instrument Clinical electrocardiography deals with the diagnostic interpretation of records obtained in man based in part on empirical correlations and in part on fundamental physiologic and physical principles *See* ELECTROPHYSIOLOGY (HEART)

Electrocardiographic leads These are records of differences in potential occurring at two specific electrode positions (*see* POTENTIAL, ELECTRIC)



Electrocardiographic tracings (a) Response characteristics of commercial electrocardiographs showing rapid deflection time and sustained response to a dc voltage (b) Normal *electrocardiogram* A small low voltage deflection is caused by atrial excitation (P wave), followed by a resting *interval (PR segment)* which denotes passage of electrical impulses from atria to ventricles The rapid tall deflection signals ventricular excitation (QRS group) the slow *deflection thereafter* ventricular recovery (T wave) A small slow deflection (V wave) occasionally follows T Its significance is uncertain (c) Abnormal electrocard *ogram* P waves are notched and separated from the ventricular complexes by a long PR interval delay in atrioventricular conduction time (d) Abnormal electro*cardiogram* P waves are replaced by sawtooth shaped undulations, ventricular complexes are irregularly spaced and QRS is widened atrial flutter with defect in *intraventricular* conduction (e) Abnormal electrocardiogram Rapid and grossly abnormal ventricular complexes without discernible atrial activity (paroxysmal) ventri*cular tachycardia* (f) Abnormal electrocardiogram Normal sequence of P, QRS, and T waves QRS is decreased in size and T the recovery deflection is deformed with the take off from the QRS group elevated acute myocardial infarction resulting from occlusion of a coronary artery

Standard bipolar limb leads refer to potential differences between two extremities when each is connected to one of the input terminals of the recorder Unipolar leads are obtained when one electrode the exploring electrode is coupled with an "indifferent" electrode whose potential is close to the mean potential of the body during the entire cardiac cycle A neutral electrode of this kind may be obtained by connecting three extremities over appropriate resistors to a common central terminal which in turn is connected to one of the input terminals Unipolar leads placed in the vicinity of cardiac muscle are said to be preferentially influenced by electrical events subjacent to the electrode position Although somewhat debatable on theoretic grounds this is a fact of clinical usefulness

The size of an electrocardiographic deflection in surface records is relatively uniform in most species and varies approximately from 0 1 to 30 millivolts (mv), with a dominant frequency range of 5–40 cycles per sec (cps) Peak voltages may be reached in 20 30 milliseconds (msec) with a total duration of the event varying from 250 to 500 msec depending on the heart rate The characteristics of an electrocardiographic instrument include a flat frequency response in the range 0 5–100 cps a deflection time of 10 msec or less and a response to a direct current (dc) voltage which deviates at 0 2 sec not more than 10% from the peak response

The deflections are arbitrarily termed P QRS T and U waves These represent the surface manifestations of excitation and recovery of atrial and ventricular musculature The interval between the end of P and the beginning of the QRS group known as the PR interval, defines atrioventricular impulse conduction time and the QT interval the total duration of the excitatory state of ventricular musculature

Clinical analysis. The clinical analysis of electrocardiograms is concerned with alterations in the form of the complexes induced by abnormalities in conduction pathways, rate of cellular polarization reversal myocardial electrolyte imbalance cardiac chamber enlargement, and pathologic changes secondary to alterations in myocardial blood supply Interpretations of changes in cycle sequence in itself a cause of variations in shape of the complexes, have laid the basis for a precise study on the nature of the irregular heart *See* BIOPHYSICS, BIOPOTENTIALS AND ELECTROPHYSIOLOGY, VECTOR CARDIOGRAPHY [H HE]

Bibliography H H Hecht, *Basic Principles of Clinical Electrocardiography*, 1950, E Lepeschkin *Modern Electrocardiography*, 1951

Electrochemical equivalent

The weight of a substance according to Faraday's law, produced or consumed by electrolysis with 100% current efficiency during the flow of a quantity of electricity equal to 1 faraday or 96 500 coulombs (1 coulomb corresponds to a current of 1

ampere during 1 second) Electrochemical equiva lents are essential in the calculation of the current efficiency of an electrode process

The electrochemical equivalent of a substance is equal to the gram atomic or gram molecular weight of this substance divided by the number of electrons involved in the electrode reaction For example The electrochemical equivalent of zinc for which two electrons are required to deposit one atom is

$$\frac{Zn}{2} \quad \text{or} \quad \frac{65\,38}{2}\,g$$

Thus the faraday is equal to the product of the charge of the electron times the number of elec trons (the Avogadro number) required to react with 1 atom or molecule equivalent of substance The value of the faraday computed in this manner agrees with values obtained from electrochemical determinations The relative error on the value 96 500 coulombs is smaller than $\pm 0\,01\%$ *See* COULOMETER ELECTROLYSIS [P D]

Electrochemical process

The principles of electrochemistry may be adapted for use in the preparation of commercially impor tant quantities of certain substances both inor gan c and organic in nature

INORGANIC PROCESSES

In rganic chemical processes can be classified as electrolytic electrothermic and miscellaneous processes including electric discharge through gases and separation by electrical means In elec trolytic processes chemical and electrical energy are interchanged Current passed through an elec trolytic cell causes chemical reactions at the elec trodes Voltaic cells convert chemicals into electric ity Electrothermic processes use electricity to attain the necessary temperature for reaction *See* ELECTROCHEMISTRY ELECTROLYSIS ELECTROLYTIC CONDUCTANCE ELECTROMOTIVE FORCE (CELLS)

For a discussion of equipment used to convert alternating current to direct current for electrolytic plants *see* CONVERTER SYNCHRONOUS MECHANI CAL RECTIFIER MERCURY VAPOR RECTIFIER MOTOR GENERATOR SET SEMICONDUCTOR RECTIFIER

Voltaic cells are used for the intermittent pro duction of small amounts of electricity When the chemicals involved are exhausted and must be re placed the unit is called a primary cell If the ex hausted components can be revived by passing electricity backwards through the unit it is called a secondary cell storage battery or accumulator Cell are connected in series to form a battery

For a discussion of the theory and description of commercial primary and secondary cells *see* BAT TERY (ELECTRIC)

Electrolysis in aqueous solutions The electroly sis of water to form hydrogen and oxygen according to the equation $2H_2O \rightarrow 2H_2 + O_2$ may be con sidered as the simplest process for aqueous elec trolytes The process is important for hydrogen production The electrolyte is a 25% solution of NaOH or 34% KOH The cathode is steel and the anode nickel plated steel Diaphragms between electrodes prevent mixing of the gases which pass separately to collecting chambers above the respec tive electrodes *See* HYDROGEN OXYGEN

Metallurgical applications Protective or decora tive coatings on base metal such as steel are ob tained by electroplating The final desired surface may require several layers of different metals or even layers of the same metal deposited under varying conditions Plating of copper cadmium chromium cobalt gold iron lead nickel platinum rhodium silver tin and zinc and of alloys such as brass are all practiced Factors affecting the re sulting plate include pretreatment and cleaning of the metal surface current density concentration of metal ions agitation temperature conductance of solution pH and addition agents *See* ELECTRO PLATING OF METALS

Electroforming is the formation of articles such as electrotype by electrodeposition A mold of nonmetal such as wax is made conducting by coat ing with powdered graphite or metal Electrodepo sition is completed and the mold is removed by melting to leave the electroformed article

In the anodizing of aluminum articles a coating approximately 0 001 in thick is applied which can be dyed or made impervious The article is cleaned and made anodic in sulfuric acid solution

Electrolytic polishing of metals is accomplished by making the article anodic in a suitable elec trolyte High points on the surface apparently dis solve to give a polish equal to or better than me chanical polishing

Electrorefining refers to purifying a metal by making it anodic in a suitable electrolyte and obtaining pure metal at the cathode The process is

in the electrolyte between anode and cathode

Electrometallurgy refers to recovery of a metal

for copper zinc cadmium and manganese *See* ELECTROMETALLURGY see also separate articles on specific metals

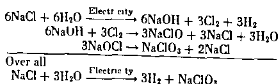
Electrolytic corrosion of metals occurs because some parts of the surface act as anodes and cor rode whereas other parts act as cathodes and do not corrode

Cathodic protection is provided if the whole sur face is made cathodic to a separate anode and sufficient voltage is available between the two elec trodes This is used to inhibit corrosion of boilers condensers underground pipelines ships and water tanks Sacrificial anodes of zinc or magne sium may provide the potential or inert anodes such as graphite stainless steel or platinum plated titanium may be used with power supplied from a rectifier *See* CORROSION

Alkali chlorine processes Electrolysis of alkali halides is the basis of the alkali chlorine and chlorate industries Chlorine Cl_2 and caustic soda NaOH (or caustic potash KOH) are made in the diaphragm cell process by interposing an asbestos diaphragm between a graphite anode and an iron screen cathode Saturated brine fed around the anode passes through the diaphragm to the cathode Chlorine is formed at the anode Hydrogen is released at the cathode leaving NaOH as a 10–15% solution and 10 15% residual salt in the brine By evaporation to 50% NaOH the salt crystallizes out and is recycled Chlorine and NaOH (or KOH) are also made by the mercury cell process by electrolyzing brine between graphite anodes and a mercury cathode forming a dilute sodium amalgam which is decomposed in another compartment by water in contact with graphite surfaces to form H_2 and NaOH

Sodium hypochlorite is formed when the products of the electrolysis of brine are mixed Electrolytic cells have been built for this purpose but sodium hypochlorite is normally made chemically

Sodium chlorate is made in cells with graphite anodes and steel cathodes where mixing occurs according to several reactions



$$6NaCl + 6H_2O \xrightarrow{\text{Electricity}} 6NaOH + 3Cl_2 + 3H_2$$
$$6NaOH + 3Cl_2 \rightarrow 3NaClO + 3NaCl + 3H_2O$$
$$3NaOCl \rightarrow NaClO_3 + 2NaCl$$

Over all

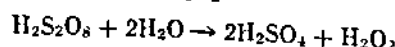
$$NaCl + 3H_2O \xrightarrow{\text{Electricity}} 3H_2 + NaClO_3$$

The temperature is kept below 40°C and hydrochloric acid is added to maintain pH 6 8 Sodium dichromate prevents reduction of chlorate at the cathode Salt is added and electrolysis is continued until the sodium chlorate has reached the desired concentration It is then recovered by crystallization See CHLORINE POTASSIUM SODIUM

Oxidations and reductions These reactions occur in all cells but in a narrower sense oxidation reactions are those in which oxygen or chlorine at the anode oxidize some material to form a new compound reduction reactions are those in which hydrogen liberated at the cathode reduces a material to a new product There are no commercial applications of inorganic electrochemical reductions

Sodium perchlorate is made by oxidation from a solution containing $NaClO_3$ at pH 6 1–6 4 and using a platinum or lead dioxide anode and an iron cathode

Persulfuric acid $H_2S_2O_8$ is made by oxidizing sulfuric acid as an intermediate in the production of hydrogen peroxide H_2O_2



$$H_2S_2O_8 + 2H_2O \rightarrow 2H_2SO_4 + H_2O_2$$

The cell has smooth platinum anodes a porous stoneware diaphragm and a lead cathode cooled to 30°C Hydrogen peroxide is recovered by distillation See PEROXIDE

Electrolytic manganese dioxide for special dry cells is made by electrolyzing hot $MnSO_4$ H_2SO_4 solutions with graphite anodes and lead cathodes

Ion permeable membrane cells These utilize diaphragms made of ion exchange resins Cation permeable membranes permit cations to pass through but not anions whereas the reverse holds for anion permeable membranes They have been applied to the purification of sea water, other applications will undoubtedly follow *See* ION PERMEABLE MEMBRANE

Fused salt electrolysis Fluorine aluminum magnesium sodium, lithium beryllium calcium, cerium and misch metal are obtained by electrolysis of fused salts because water interferes with the desired reaction Raw materials must all be purified before addition to fused salt cells because purification of the electrolyte is not economical as in aqueous electrolytes

Fluorine is produced by electrolysis of 40% HF in KF between carbon anodes and steel cathodes at 100–110°C. A diaphragm of Monel screen keeps the products H_2 and F_2 separated Dry HF gas is bubbled continuously into the electrolyte

Aluminum is produced in carbon lined steel pots containing an electrolyte of alumina dissolved in fused cryolite (AlF_3 3NaF) at 950–1000°C. The anode is also carbon The pool of aluminum in the bottom of the pot is the cathode It is siphoned out periodically Oxygen released at the anode reacts with the carbon to form carbon monoxide

Magnesium is produced by electrolysis of fused 25% $MgCl_2$ 75% NaCl at around 700°C. The Dow process for making magnesium from sea water uses cell feed material approximating $MgCl_2$ $2H_2O$ which is fed around the graphite anodes where dehydration occurs Gas from the anode compartment is wet chlorine air, and hydrogen chloride which is used to make fresh magnesium chloride from magnesium hydroxide Magnesium metal is deposited on steel cathodes which direct the metal to a collecting zone The cell is a cast steel pot in a furnace setting European cells use molten anhydrous magnesium chloride feed They have brick lined steel bodies with graphite anodes Concentrated chlorine from the cells chlorinates MgO and coke in electrically heated shaft furnaces from which molten $MgCl_2$ is tapped periodically to feed the cells Molten magnesium is ladled from the cells and cast into molds protected by an atmosphere of SO_2

Sodium was once made by electrolysis of fused NaOH but since 1929 has been made by electrolysis of NaCl in the Downs cell The electrolyte is sodium chloride–calcium chloride eutectic (33 2% NaCl) at 600°C The cell consists of a brick lined steel vessel Graphite anodes project upward from the bottom The cathode is made of steel cylinders concentric with the anodes and supported from iron arms extending through the sides of the cell A diaphragm of steel screen directs the sodium into an inverted trough leading to a riser pipe which conducts it to a collecting tank above the cell Chlorine is collected in an inverted cone over the anode Pure dry salt is fed to the cell

Lithium is made in a cell similar to the Downs sodium cell except that the electrolyte is 60% LiCl 40% KCl at 450–500°C.

Calcium has been made by electrolysis of pure fused calcium chloride at about 800°C In this case the cathode is solid calcium and is mechanically withdrawn from the cell as a "carrot"

Beryllium is made by batch electrolysis of fused salt starting with 25% BeCl and 75% NaCl in a chrome-iron pot acting as cathode and a graphite anode At the conclusion of electrolysis the beryl lium is found clinging to the wall of the pot This is cleaned out and broken up when cold Salt is washed out with water The metal is in the form of bright crystalline plates A beryllium copper eutectic can be made by using a copper cathode

Beryllium alloys containing copper and nickel are made from BeO in arc furnaces

Cerium and misch metal are made from $CeCl_3$ or mixtures of chlorides of cerium lanthanum and neodymium in a fused salt electrolysis with NaCl as in beryllium electrolysis Misch metal is used for lighter flints

Electrothermics The manufacture of many products requires temperatures higher than can be obtained by combustion methods Electric heat can usually be developed at or close to the point where it is required

Electric furnaces are used for melting and smelting iron and steel, making ferroalloys making nonferrous metals and alloys preparing nonmetallic products such as calcium cyanamide silicon carbide graphite boron carbide fused alumina carbon disulfide phosphorus and chlorides of magnesium boron zirconium and titanium and fixing nitrogen

Methods of electric heating utilize resistance arcs or induction Resistance furnaces may use the substance being heated as a resistor or auxiliary resistors may be used Arc furnaces may have an arc between an electrode and the substance being heated or the arc may be between two or more electrodes The arc is formed by contacting the electrodes and then drawing them apart If the gap becomes too wide the arc will break A means of regulating the distance between electrodes must be provided The arc gives the highest temperature tool available The temperature is limited by the materials of the electrodes and the furnace lining Induction furnaces use the crucible or its charge as the secondary circuit of a transformer at low or high frequency alternating current *See* HEATING ELECTRIC

Processes in gases Electrical discharge through gases has industrial application in ozone production and nitrogen fixation *See* ELECTRICAL CONDUCTION IN GASES

Ozonizers consist of two metal electrodes with an air gap and dielectric such as glass between them One commercial unit operates at 15 000 volts 60 cycles sec 35–40 watts/ft² Very dry air passing through the air gap leaves containing 10–12 mg ozone per liter *See* OZONE

Fixation of nitrogen by passing air through an arc furnace and forming oxides of nitrogen is practiced where power is cheap in Norway France and Italy *See* NITROGEN FIXATION

Electric separations Magnetic separation removes tramp iron from solids suspensions and solutions Magnetic separators are also used to separate solids of various magnetic susceptibilities

The Cottrell electrostatic precipitator removes dusts and mists from gases In one form a fine wire is axial in a pipe and insulated from it The pipe is grounded and the wire is negative in a high voltage direct current circuit Particles in gases passing through the pipe become electrically charged and move to the pipe Liquid particles drain off while solids are periodically vibrated off

Electrophoresis is the migration of colloidal particles which acquire positive or negative charges in an electric field Rubber particles in latex have a negative charge and may be deposited on an anode This is used for the application of rubber coatings to odd shaped metal articles

Electroendosmosis is the migration of water through a porous diaphragm in an electric field This has been applied to dewatering of clays and peat *See* ELECTROOSMOSIS ELECTROPHORESIS ELECTROSTATIC PRECIPITATOR SEPARATION (MECHANICAL) [W C G]

ORGANIC PROCESSES

Most electroorganic processes involve anodic oxidation or cathodic reduction of an organic compound at insoluble metal electrodes the organic compound usually being dissolved in an electrically conductive electrolyte Occasionally a soluble metal electrode is used as for the production of metal organic compounds A few products have been produced by chemical action in an electrical discharge *See* ELECTRICAL CONDUCTION IN GASES ELECTROLYSIS

Products Organic chemical products that have been prepared by electrolytic methods in the laboratory are large in number but only a few processes have been carried out on an industrial scale The latter include oxidation of glucose to gluconic acid and anthracene to anthraquinone and reduction of glucose to mannitol and sorbitol azobenzene to benzidine fumaric acid to succinic acid and pyridine to piperidine Technical scale production was encouraged in the early 1900s by successful production of dyestuff intermediates by electroreduction of nitro compounds Later in 1949 the difficult problem of fluorination of organic compounds was piloted as an electrolytic process The desalting of sugar cane juice with electrolytic ion exchange membranes was announced in 1958

Beneficiation of natural gas (as in the conversion of methane to acetylene benzene olefins or higher hydrocarbons) is an example of the application to chemical action in electrical discharge Oils of low viscosity have been converted to lubricating grade by passing a discharge through thin continuously renewed layers until the desired vis

cosity is attained Commercialization of reactions in electrical discharge has been rare however Current efficiencies exceed Faraday s law but the high voltages required result in low energy yields and the reactions produce a diversity of products which are difficult to separate

Large tonnages of organic chemical products are made with materials manufactured by inorganic electrochemical processes The close relationship of the caustic chlorine and electrothermic industries to electroorganic chemistry is illustrated by the product listing in the table

Products derived from inorganic electroprocesses

Organic compound produced	Inorganic electroprocess product used
Glycols	Chlorine
Trichloroethylene	Chlorine
Vinyl chloride	Chlorine
Insecticides (DDT p-dichlorobenzene)	Chlorine and caustic soda
Textiles	Chlorine and caustic soda
Petroleum refining	Caustic soda
Soap cleansers, cleaners	Caustic soda
Tetraethyllead	Chlorine sodium
Sulfa drugs cyanides organic synthesis	Calcium cyanamide
Acetylene	Calcium carbide
Fatty alcohols	Sodium dispersions and sodium amalgams

Applications The applicability of electroorganic synthesis on a commercial scale is determined in part by cell capacity or floor space solubility of the organic reactant power cost and complexity and expense of equipment and control Where possible desirable oxidation and reduction reactions should be carried out simultaneously in the same cell to minimize energy costs The simultaneous oxidation of anthracene to anthraquinone and reduction of azobenzene to benzidine has been achieved in a cell of commercial size

The technology of electroorganic chemistry has been extended to areas beyond the immediate sphere of preparative chemistry Th[illegible]

[illegible] of the solution using soluble aluminum anodes and to engineering quality aluminum plate from an ether [illegible]

[illegible] from electroorganic reactions Since about 1952 interest in the development of batteries containing organic depolarizers (active materials) has revived Their [illegible]

[illegible] chemicals by electrolysis are cell design current density, and temperature Diaphragm cells are used ordinarily to restrict the organic compound being converted (the depolarizer) to the vicinity of the electrode at which it is to be oxidized or reduced and to permit separate control over the acidity and composition of anolyte and catholyte Electrolytic reduction is in general, more rapid and complete and depolarization efficiency is higher when the current density is low High depolarizer concentration and effective stirring permit the effective use of higher current densities High operating temperatures also increase the rate of reduction of the depolarizer The limited solubility of many organic compounds in conductive aqueous electrolytes of desirable acidity is a major problem in connection with high rate of production Depolarizers of limited solubility may be brought into contact with the electrode as suspensions or emulsions The area of contact of suspended or emulsified materials with the electrode may be increased by use of porous electrodes Solid depolarizers may even be incorporated in the electrode itself as a paste of depolarizer and electrode material

Typical reactions. Organic oxidation reduction systems are seldom reversible With the exception of a few systems of biological interest the only ones which have so far been shown to be thermodynamically reversible are quinones, quinoimines quinone diamines, azobenzenes and nitrosobenzenes, indophenol indamine azine oxazine thiazine indigoid and flavin dyes arylarsonic acids and nitroguanidine, and stable free-radical ions such as triphenyl and methyl ion These relatively few organic compounds can be converted reversibly to their normal reduction products if they are present as cathode depolarizers in the catholyte of an electrolytic cell or their reduction products can be reversibly oxidized if present as depolarizer material in the anolyte of an electrolytic cell It is apparent from this that, in addition to being electrolysis products these compounds could be used as depolarizers in batteries

The ease and speed of electrolytic oxidation and reduction of the above compounds depend mostly on the acidity of the electrolyte preferably alkaline for oxygen type compounds such as quinone and acidic for nitrogen type compounds such as quinone-diamines The acidity determines the stability of intermediate reaction products formed by electrode reactions involving only one electron such as quinone to semiquinone Whenever a bivalent oxidation can be brought about by two successive univalent oxidations the speed of such a reaction will be greatly enhanced A counterpart in inorganic chemistry is the rapid oxidation of Tl^{+} to Tl^{3+} by Ce^{4+} in the presence of manganese salts which can lose one two or three electrons at a time in being reduced from Mn^{+} through Mn^{+} and Mn^{2+}

Effect of electrode type. Catalytic influences on the electrode reactions for electroorganic preparations involve the use of the electron carrier ions in salts of chromium titanium tin cerium vanadium and iron These ions of variable oxidation number react readily with the organic substance and then are regenerated at one of the electrodes For example, the efficiency of the oxidation of

terpene hydrocarbons at a lead dioxide anode is increased by the addition of a small amount of a ceric salt to the electrolyte very likely because the cyclic reduction of ceric ion by the hydrocarbon and preferential oxidation of cerous ion at the anode The metal electrode can also have a catalytic influence on the reactions Oleic acid is reduced more smoothly to stearic acid at palladium or platinum cathodes than at nickel cathodes Reduction products of nitrobenzene in alcoholic solution are azoxybenzene and azobenzene at platinum nickel or mercury cathodes whereas the principal reduction product is aniline at a copper cathode

Electrode potential partly determines the course of electroorganic reactions The need for carefully controlled electrode potential is an economic factor in electroorganic processes Electrodes with high overpotentials contribute to high reducing or oxidizing energies A cathode of mercury or lead is required for reduction of pyridine and compounds containing the keto group Mercury amalgamated lead or some other metal of high hydrogen over potential is required to reduce hexoses to their corresponding hexatols A few low overpotential metals notably copper and palladium are singularly effective cathodes Copper is as good as mercury for the reduction of sorbic acid and is an effective cathode material for the reduction of nitro compounds The reduction of oleic to stearic acid occurs smoothly and in good yield at palladium cathodes

High purity electrodes are essential to good efficiency for difficult reductions Traces of low over potential impurities should also be absent from the electrolyte The reduction of caffeine for example is completely inhibited when traces of salts of copper silver or platinum are contained in the electrolyte Likewise dissolved iron salts reduce the current efficiency of pinacol formation in the reduction of acetone at a lead coated copper cathode

The physical condition of electrodes also influences their efficiency The efficiency of reduction of n propyl ketone to pentane at cadmium amalgam electrodes varies with phase changes in the alloy In the electrolytic reduction of hexoses a cast zinc cathode (001 face of the crystal parallel to surface) is better than an electrodeposited zinc cathode (110 face of the crystal parallel to surface) the activity of a lead cathode increases with increase in grain size Liquid metal cathodes used for reduction of sorbic acid to salts of dihydro acids in alkaline solution produce appreciable amounts of a bimolecular product which is obtained in much lower yield at solid cathodes *See* OXIDATION REDUCTION

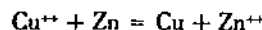
[L D M]

Bibliography H J Creighton and W A Koehler *Principles and Applications of Electrochemistry* vol 2 2d ed, 1944 C L Faust Modern industries depend on electroprocesses *J Electrochem Soc* 98(10) 133C 1951, C L Mantell *Industrial Electrochemistry*, 3d ed 1950

Electrochemical series

A series in which the metals are listed in the order of their chemical reactivity, the most active at the top and the less reactive or more "noble" metals at the bottom In a broader sense such an activity series need not be limited to the metals but may be carried on through the electronegative (nonmetallic) elements as well (See table for list of common elements)

The electrochemical series as it applies to metals was first established by laboratory experiments in which the purpose was to determine which metals would displace others from solutions of their salts Thus a clean strip of zinc immersed in a solution of copper sulfate is soon found to be covered by a deposit of copper while zinc in turn goes into solution from the strip as zinc ions By definition then zinc is a more reactive metal than copper, since it will displace copper from a solution of Cu^{++} ions The reaction is readily seen to be an oxidation reduction transfer of electrons which can be summarized by the equation



$$Cu^{++} + Zn = Cu + Zn^{++}$$

Similarly copper will displace silver from a solution containing Ag^{+} ions depositing crystals of metallic silver and coloring the solution with Cu^{++} ions From these observations an activity series may be set up in the order Zn Cu Ag By exhaustive experiments with other metals it becomes possible to draw up a complete list in the order of chemical activity in which the metals at the top of the list are those which are found to give up their electrons most readily (that is are the most electropositive elements)

The ease with which an isolated atom of an element gives up an electron known as the first ionization potential is a precise physical quantity which can be measured by electrical experiments on gases or vapors at low pressure (*see* IONIZATION POTENTIAL) The replacement experiments which determine the order of the electrochemical series take place in a very different environment since they involve solid phases and also aqueous solutions with their consequent hydration effects Moreover it might well be expected that displacement reactions in solution would depend upon the concentrations of the reagents used and also upon the presence or absence of other dissolved substances To obtain a more accurate and reproducible activity series it is best to turn to the more exact quantity called electrode potential or oxidation reduction potential which is defined as the voltage developed by a sample of pure metal immersed in a solution of one of its salts (at unit activity and at 25°C) versus a hydrogen electrode immersed in hydrochloric or sulfuric acid of equivalent concentration For further details about this measurement *see* ELECTRODE POTENTIAL It is evident that by confining the experimental conditions to a standard concentration and temperature the hydration and concentration effects may be kept

Electrochemical series of the elements*

Lithium	Li	Aluminum	Al	Molybdenum	Mo
Potassium	K	Titanium	Ti	Tin	Sn
Rubidium	Rb	Zirconium	Zr	Lead	Pb
Cesium	Cs	Manganese	Mn	Germanium	Ge
Radium	Ra	Vanadium	V	Tungsten	W
Barium	Ba	Niobium	Nb	Hydrogen	H
Strontium	Sr	Boron	B	Copper	Cu
Calcium	Ca	Silicon	Si	Mercury	Hg
Sodium	Na	Tantalum	Ta	Silver	Ag
Lanthanum	La	Zinc	Zn	*Gold*	Au
Cerium	Ce	Chromium	Cr	Rhodium	Rh
Magnesium	Mg	Gallium	Ga	Platinum	Pt
Scandium	Sc	Iron	Fe	Palladium	Pd
Plutonium	Pu	*Cadmium*	Cd	Bromine	Br
Thorium	Th	Indium	In	Chlorine	Cl
Beryllium	Be	Thallium	Tl	Oxygen	O
Uranium	U	Cobalt	Co	Fluorine	F
Hafnium	Hf	Nickel	Ni		

* According to standard oxidation potentials E^0 at 25°C

quite constant making possible a more exact listing of metals according to their activity. Hence any present day electrochemical series must rely on the measurements of oxidation potential and should be in agreement with the accepted values determined from such electrochemical cells. Such reliance has the further advantage that the series need not then be confined to metals but may be extended to the nonmetals or electronegative elements. As before those metals which will liberate hydrogen from dilute acids (such as hydrochloric or sulfuric) will stand above hydrogen in the series and will have positive oxidation potentials while those metals and nonmetallic elements which will not liberate hydrogen from such dilute acids will stand below hydrogen in the list and will have negative oxidation potentials. Since the oxidation potentials are also related to the equilibrium constants for reversible reactions it becomes possible to calculate oxidation potentials from other information where direct experiments are inconvenient, as in the case of the alkali metals versus aqueous solutions of their salts. See ELECTROCHEMISTRY, ELECTRONEGATIVITY, OXIDATION REDUCTION.
[E C R]

Bibliography W M Latimer, *The Oxidation States of the Elements and Their Potentials in Aqueous Solutions*, 2d ed , 1952

Electrochemistry

That portion of physical chemistry which includes the phenomena which occur when electric current passes through electrolytes, or through junctions between electrolytes and metallic conductors. This includes the reactions that take place in galvanic cells that is, electric batteries.

Since all matter is fundamentally electrical all chemistry is essentially electrochemistry. However, in general usage, the term has a much more restricted meaning which can be outlined by the consideration of electrically conducting systems. These are of three types: metals, electrolytes, and gases though there are intermediate kinds, such as semiconductors.

Conduction in metals is due to electrons which are subatomic in size. In electrolytes conduction derives from the movement of ions which are electrically charged atoms, molecules, or molecular aggregates. Typical electrolytes are fused salts and aqueous solutions of salts, acids, and bases. A large part of electrochemistry deals with the properties of electrolytes (see ELECTROLYTIC CONDUCTANCE). This part of the subject involves identification of the ions, their number per unit volume and their speeds, or mobilities, in gradients of electric potential. The investigation of conduction in metals and in gases, is generally considered a portion of physics.

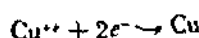
Branches of electrochemistry. A most important domain of electrochemistry is the study of the phenomena which occur when electric current passes from metallic to electrolytic conductors or vice versa. At such surfaces electrochemical reactions must occur. If the electrochemical reaction yields electrons to the metal the surface is called an anode, and the reaction is oxidizing in nature. Conversely, when the metal supplies the electrons the surface is a cathode, and the electrochemical reaction is reducing.

Nearly all the electrochemical systems of scientific or industrial interest involve two metallically conducting electrodes, with one or more electrolytic conductors between them. In many important cases the reaction at one electrode when current is passed is the reverse of that at the other electrode. In copper refining, for instance, the reaction



$$Cu \rightarrow Cu^{++} + 2e$$

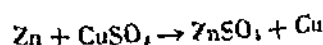
(in which e represents the electron) takes place at the anode and



$$Cu^{++} + 2e^- \rightarrow Cu$$

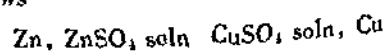
at the cathode. Many electroplating processes are of this kind.

A system in which the reaction at the anode differs from that at the cathode is in essence, a galvanic cell. Primary cells, that is batteries are of this type and are the means by which certain chemical reactions may be made to yield electrical energy. Thus for instance, if metallic zinc is placed in a copper sulfate solution the reaction



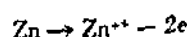
$$Zn + CuSO_4 \rightarrow ZnSO_4 + Cu$$

occurs. The same reaction may be made to take place in a galvanic (Daniell) cell which is arranged as follows



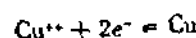
$$Zn, ZnSO_4 \text{ soln } CuSO_4 \text{ soln}, Cu$$

which gives a potential of about 1 volt. The electrochemical reaction at the anode is



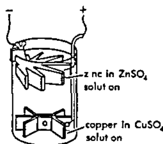
$$Zn \rightarrow Zn^{++} - 2e$$

and at the cathode



$$Cu^{++} + 2e^- = Cu$$

A type of Daniell primary cell which was much



The Don ell cell

u ed is the gravity cell shown in the illustration The Cu cathode is surrounded by saturated $CuSO_4$ solution on which is floated a $ZnSO_4$ solution with the Zn anode

Though they cannot be separately measured elec tr c potential are presumed to exist at the contacts between electrolytic conductors An example is the contact between the zinc and copper sulfate solu tions in the Daniell cell as given above Such liquid junctions are especially important in biological phenomena

Galvanic cells are of utility in many types of sci entific investigation Certain cells may be used to determine the free energies of the reactions taking place in them They are of service in the study of ox dation reduction potentials and in the determi nat on of ionization constants Many analytical methods depend upon the measurement of the po- tentials of galvanic cells as does the determination of alkalinity and acidity or pH values Other gal vanic cells may be used as sources of limited amounts of electric power One of the most impor tant is the Leclanche dry cell which consists of a zinc anode and a carbon cathode surrounded by a mixture of powdered carbon and manganese diox ide The electrolyte is ammonium chloride

Galvanic cells may also be used for the tempo rary storage of electrical energy The most impor tant of these cells is the lead storage battery or lead accumulator This consists of a lead peroxide electrode and a lead electrode with sulfuric acid as electrolyte When furnishing current both elec trodes change to lead sulfate and when recharged by rever ing the direction of the current the elec trodes return to their original composition Other storage cells have been produced but so far have but lim ted use

Less common phenomena A number of electro chemical phenomena are grouped together as elec trokinetic phenomena The displacement produced by an applied electromotive force of a liquid with reference to the surface of a solid is termed electro- osmos s Electrophoresis (or cataphoresis) is the movement of suspended solid colloidal or liquid particles in an electric field Particularly as ap plied to proteins and related substances electro phoresis has found wide application in biological research The electromotive force produced by forcing a liquid through a capillary tube or a po- rous plug is the streaming potential

The important topic of dielectric constants of solids and liquids is usually considered to be a por tion of physics However the effect of the structure of the molecules on these constants has electro chemical interest as has the distribution of electric charges on the molecules which results in the pres ence of dipole moments

Some metals notably iron and chromium can exist in active and passive states because of the presence of surface films In the latter state pro duced by strong oxidizing agents they are rela tively insoluble in acids The active state may be restored by abrasion or contact with active metals

The overvoltage (or overpotential) of an electro chemical reaction is the difference between the potential of an electrode at which the reaction is actively taking place and another electrode which is at the equilibrium condition for the same reac tion Large overvoltages are observed only for the evolution of gases particularly of hydrogen or oxy gen *See* ELECTROCHEMICAL PROCESS ELECTRODE POTENTIAL ELECTROLYSIS ELECTROPLATING OF METALS [D A M]

Bibliography S Glasstone *Introduction to Elec trochemistry* 1942 H S Harned and B B Owen *Physical Chemistry of Electrolytic Solutions* 3d ed 1958

Electrode

An electrical conductor through which an electric current enters or leaves a conducting medium whether it be an electrolytic solution solid molten mass gas or vacuum For electrolytic solutions many solids and molten masses an electrode is an electric conductor at the surface of which a change occurs from conduction by electrons to conduction by ions For gases and vacuum the electrodes merely serve to conduct electricity to and from the medium *See* ELECTRODE POTENTIAL ELECTRO DEPOSITION ANALYSIS ELECTROLYSIS ELECTROMO TIVE FORCE (CELLS) [W J H]

Electrode potential

The potential which a metal or gas electrode takes up relative to a solution of ions

Metal–metal ion electrodes When a metal is immersed in an electrolyte an equilibrium tends to be established in which a steady difference of electric potential exists across the region of the

ionize If the metal electrode represented by M has a valence of n the reaction which takes place is then

$$M \rightleftharpoons M^{n+} + ne$$

where e^- indicates an electron and M^{n+} an ion in solution Examples of metal metal ion electrodes of this type are zinc copper and sodium e



Because the potential of such an electrode changes with the concentration of ions it is necessary to adopt some standard concentration at which to compare the potentials of various electrodes. The standard electrode potential E expressed in volts is defined as the potential of an element immersed in a solution of its ions at unit activity, that is, the effective concentration of 1 mole/1000 g of water. The electrode potential E at other concentrations is given by the expression

$$E_M = E^0_M - \frac{RT}{nF} \ln a_{M^{n+}}$$

where T is the absolute temperature, F is the Faraday (96 484 coulombs), R the gas constant, and $a_{M^{n+}}$ the effective concentration (activity) of M^{n+} ions in the solution. At 25°C the above expression can be written as

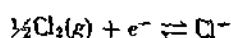
$$E_M = E^0_M - \frac{0.05916}{n} \log a_{M^{n+}}$$

Because the single electrode potential E involves the activity of an individual ionic species, it has no strict thermodynamic significance. This difficulty is overcome by defining the standard hydrogen electrode as an arbitrary zero of potential. Electrode potentials based on this zero are thus said to refer to the hydrogen scale. Such a potential is actually the emf of a cell obtained by combining the given electrode with a standard hydrogen electrode. See ELECTROCHEMICAL SERIES; HYDROGEN ELECTRODE.

The various electrodes encountered in electrochemical work may be grouped into seven types: (1) metal-metal ion, (2) amalgam, (3) nonmetal-nongas, (4) gas, (5) metal-insoluble salt, (6) metal-insoluble oxide, and (7) oxidation-reduction. Any of these electrodes may be combined with any other to give a cell, the electromotive force of which is equal to the algebraic sum of the potentials of the two electrodes. Metal-metal ion electrodes and gas electrodes are discussed in this article.

Gas electrodes. A gas electrode is formed by partially immersing an inert metal (usually platinized platinum) in a solution of the ions of the gas. The gas must establish a reversible equilibrium with the ions in solution in the presence of the metal. The function of the metal wire or foil is to facilitate establishment of equilibrium between the gas and its ions and to serve as the electric contact for the electrode.

The potential of such an electrode is determined by the pressure of the gas and the activity of its ions in solution. Thus for the chlorine electrode the reaction is



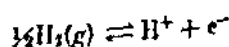
$$\tfrac{1}{2}Cl_2(g) + e^- \rightleftharpoons Cl^-$$

and the equation for the electrode potential E_{Cl_2} is

$$E_{Cl_2} = E^0_{Cl_2} - \frac{RT}{F} \ln \frac{a_{Cl^-}}{P^{1/2}_{Cl_2}}$$

where P_{Cl_2} is the pressure of chlorine in atmospheres, a_{Cl^-} the effective concentration (activity) of chloride ions in the solution, and $E^0_{Cl_2}$ the standard electrode potential for the chlorine electrode, which is equal to 1.3583 volts at 25°C. The standard electrode potential of a gas electrode is defined as the potential of the electrode when the gases involved in the reaction are at a fugacity of 1 atm, that is, an effective pressure of 1 atm, and all dissolved substances are at an effective concentration (activity) of 1 molal, that is, 1 mole/1000 g of water. See ACTIVITY (THERMODYNAMICS); FUGACITY.

The most important gas electrode is the hydrogen electrode which is reversible to hydrogen ions. The reaction for this electrode is



$$\tfrac{1}{2}H_2(g) \rightleftharpoons H^+ + e^-$$

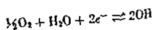
and the equation for the electrode potential given by

$$E_{H_2} = E^0_{H_2} - \frac{RT}{F} \ln \frac{a_{H^+}}{P^{1/2}_{H_2}}$$

But $E^0_{H_2}$, the standard electrode potential of hydrogen, is the reference of all emf measurements and is taken by definition to be zero at all temperatures. Thus the above equation becomes

$$E_{H_2} = -\frac{RT}{F} \ln \frac{a_{H^+}}{P^{1/2}_{H_2}}$$

Another gas electrode which has received considerable attention is the oxygen electrode, whose potential depends on the activity of hydroxyl ion. However, unlike the hydrogen and chlorine electrodes, the oxygen electrode cannot be made reversible because no suitable electrode material has been found which can catalyze the establishment of the equilibrium between oxygen and hydroxyl ions

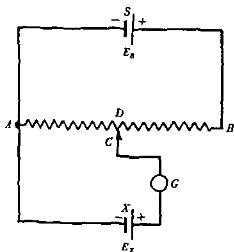


$$\tfrac{1}{2}O_2 + H_2O + 2e^- \rightleftharpoons 2OH^-$$

The standard potential of the oxygen electrode cannot be determined directly from emf measurements because of the irreversible behavior of this electrode. It is possible, however, to derive the value in an indirect manner, and it has been found to be +0.401 volt.

Electrode potential measurements. In order to measure the potential of any electrode it is necessary in principle to combine the electrode with a ... hydrogen electrode ... ential and to ... no half-cell ...

The potential of the reference electrode is then subtracted to give the required electrode potential. For various reasons, such as the difficulty in setting



Poggendorff compensation method for measuring emf

up a hydrogen gas electrode, several subsidiary reference electrodes, whose potentials are known on the hydrogen scale have been devised The most common of these is the calomel electrode which consists of mercury in contact with a solution of potassium chloride saturated with mercurous chloride *See* CALOMEL ELECTRODE, SILVER CHLORIDE ELECTRODE

A simple voltmeter cannot be used alone for measuring the emf of a small cell because the operation of the voltmeter draws some current which causes chemical changes at the electrodes and produces a different voltage To avoid these difficulties the emf is measured by balancing [illegible]

with S in such a way that the two emfs oppose each other is the source X of unknown potential E_x To find E_x the sliding contact C is moved along AB until a position D is found at which the galvanometer G gives no deflection From E_S and the distances AB and AD, the unknown emf E_x is found as follows Since E_S is impressed across the full length AB E_S for any given current passing through the resistance must be proportional to AB Again because E_x is impressed only across the distance AD, it must be proportional to this length Consequently, dividing E_x by E_S gives

$$\frac{E_x}{E_S} = \frac{AD}{AB} \quad \text{and} \quad E_x = \left(\frac{AD}{AB}\right) E_S$$

At the present time the majority of emf measurements are made by means of special potentiometers available commercially, which operate on the foregoing principle In these, the conductor AB consists of a number of resistance coils with a movable contact, together with a slide wire for fine adjustment A standard cell is used for calibration purposes and the emf of the cell being studied can then be read with an accuracy of 0 1 millivolt or better *See* POTENTIOMETER (VOLTAGE MEASUREMENT)

The standard cell that is widely employed for emf measurements is some form of the Weston cell It is highly reproducible its emf remains constant over long periods of time and it has a small temperature coefficient In order to retain constancy of emf during use only very minute currents should be drawn from the cell as is actually done if the potentiometer is operated properly *See* ELECTROCHEMISTRY, ELECTROMOTIVE FORCE (CELLS), ELECTRONEGATIVITY GLASS ELECTRODE, OXIDATION REDUCTION [R G]

Bibliography S Glasstone *Introduction to Electrochemistry*, 1942, G F A Kortum and J O'M Bockris *Textbook of Electrochemistry*, 1951, E C Potter *Electrochemistry*, 1956

Electrodeposition analysis

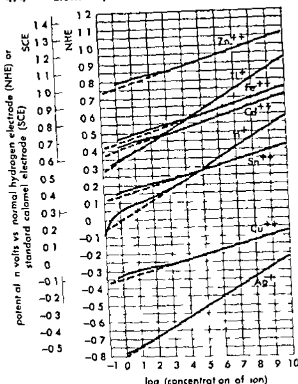
That area of quantitative chemical analysis which involves the separation of substances by electrolysis The term electrodeposition is somewhat misleading in that only separation is implied In every case the analysis must be terminated by weighing the deposited material (electrogravimetry) or by use of other physical or chemical measurements In this sense then electrodeposition is a separation method as are distillation precipitation and extraction *See* ELECTROLYSIS

This method is applied most commonly to the separation of metals from solution and from each other and their quantitative determination The quantities of metals can be determined by weighing a suitable electrode before and after deposition Among the elements that may be determined by deposition on the cathode are cadmium, nickel [illegible] silver and gold [illegible] be deposited on a silver anode as the respective silver halides

The potential of an electrode in equilibrium (no net electrodeposition occurring) with a solution of its ions of given concentration may be determined using the Nernst equation

$$E = E^0 + \frac{RT}{nF} \ln [M] \qquad (1)$$

which relates the standard potential of an electrode to the equilibrium potential in the presence of a solution of its ions In this equation, E is the equilibrium potential (versus the normal hydrogen electrode) of the electrode E^0 is the standard oxidation potential n is the number of electrons needed for the reduction of one metal ion F is the faraday (96 500 coulombs), R is the gas constant T is the absolute temperature and $[M]$, to a first approximation is the molar concentration of metal



Equilibrium electrode potentials (G W Ewing Instrumental Methods of Chemical Analysis McGraw Hill 1954)

ions in equilibrium with the electrode *See* ELECTRODE POTENTIAL

When the electrode potential is made more cathodic than the equilibrium value electrodeposition occurs until the metal ion concentration is lowered to that value which is in equilibrium with the electrode at the applied potential A tenfold change in concentration at 30°C may be effected by a 0 060/n volt change in potential By making the electrode sufficiently cathodic the metal ions remaining in the solution may be reduced to a negligible concentration Two metallic species may be separated by adjusting the potential of the cathode so that it is less cathodic than the equilibrium potential of the metal to be left in solution and more cathodic than the initial equilibrium potential of the metal to be removed Practically all of the latter metal will be deposited as equilibrium is reestablished In comparatively few cases are the eq ll m n tent l f h

convert most of them to a complex ion species After selectively complexing the metal to remain in solution the potential required to deposit the low concentration of free ions of this metal is sufficiently more cathodic than that required for the other metal to make complete separation possible

Electroseparations may be performed at the anode in cases where insoluble higher oxides are formed Equation (1) may be used in a manner similar to that described for cathodic separations, except that the reciprocal of the metal ion concentration is used in the logarithmic term along with the appropriate power of the hydrogen ion concentration As the metal ion concentration decreases the equilibrium potential of this electrode becomes more anodic

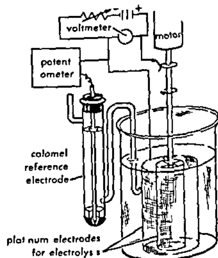
The electrode material most commonly used for electroseparations is platinum This metal is used as anode for most oxidations, and as cathode for depositions of the more readily reduced metals Because of its low overvoltage on platinum hydrogen is evolved and thereby interferes with the deposition of metals requiring very cathodic potentials for reduction

Mercury is frequently used as a cathode material because its extremely high hydrogen overvoltage permits most metals to be deposited without interference from hydrogen evolution Mercury can be removed from the deposited metals by distillation As an anode mercury finds very limited use because it is oxidized quite readily

Constant potential electrolysis The potential of the working electrode (the cathode when a metal is being deposited) is the most critical factor in obtaining a desired separation It is not possible however to obtain constant potential at an electrode by applying a constant voltage to the cell The applied voltage is expended in accordance with Eq (2)

$$E_{ap} = E_{an} - E_c + \omega_{an} + \omega_c + IR \qquad (2)$$

In this expression E_{ap} is the voltage applied to the cell E_{an} is the equilibrium potential of the anode E_c is the equilibrium potential of the cathode ω_{an} is the anodic overvoltage or increase in anode potential beyond the equilibrium value needed to pass the current of I amperes ω_c is the corresponding cathodic overvoltage and R is the ohmic resistance of the solution The overvoltages fo



Electrodeposition apparatus (G W Ewing Instrumental Methods of Chemical Analysis McGraw Hill 1954)

metal deposition and dissolution are usually less than 0.1 volt with notable exceptions for iron, cobalt, and nickel. Because they depend upon solution conditions and electrode form in a manner which is not fully understood, overvoltages cannot be predicted with certainty, thus limiting the accuracy with which the current in a given electrolysis may be calculated.

In electrolysis at constant applied voltage, the potential of a working cathode is less than the total applied voltage. When the applied voltage is adjusted so that the critical potential for a separation cannot be exceeded, the potential of the electrode is significantly less than this critical value throughout most of the deposition due to *IR* drop, and the electrolysis proceeds very slowly. If a stable reference electrode is placed in the solution and the voltage between this and the working electrode is maintained constant at the critical value by periodic adjustment of the voltage applied to the cell, much more rapid deposition results. Unfortunately, such manual control is tedious. Involved circuitry is required for the automatic control of potential. Instruments which perform this task are known as potentiostats, and several types have been described in the literature. When the cathode potential is to be controlled, the desired voltage between the cathode and reference is maintained constant by an electronic or electromechanical servo system which changes the total voltage applied between the anode and cathode. When the anode potential is controlled, the voltage between the anode and the reference is kept constant by the same means. Anode and cathode potentials cannot both be controlled at the same time.

Internal electrolysis. Electrodeposition without externally applied voltage was an early but simple method of approximating a controlled cathode potential. An active metal such as magnesium which dissolves spontaneously is made the anode of a cell and an inert electrode such as platinum is made the cathode. When the electrodes are shorted together, the potential of the cathode is equal to and created by the potential of the anode. By judicious choice of the anode metal and the concentration of the reagent in which it dissolves, the cathode potential may be made to assume predetermined values over most of the useful range. Deposition at the cathode occurs at the expense of dissolution of the anode. The current flow is limited by the magnitude of the spontaneous voltage of the cell and by its internal resistance. Internal electrolysis has been used relatively little in recent years.

Constant current electrolysis. This process precludes the possibility of electrode potential control by electrical means. While the concentration of metal ions remains large, the potential of the cathode stays near the equilibrium potential. However, as the ions in the solution are depleted, the desired cathode reaction is not able to use all of the current forced through the cell, so the potential becomes more cathodic until an additional action such as the deposition of another metal or the evolution of hydrogen takes place to maintain the current. In most electroseparation devices a large voltage is applied to the cell. In order to effect a clean separation of two metals under these conditions, it is necessary to interpose a harmless reaction to limit the potential of the cathode before it exceeds the equilibrium potential of the metal to be left in solution. The reaction most often employed is hydrogen evolution, which may be made to occur at various potentials by proper adjustment of the pH. Through use of selective complexing agents to change the relative effective concentrations of the ions (and indirectly their equilibrium potentials) and also through control of pH to limit the cathode potential, many simple combinations of metals can be separated by this form of electrolysis. *See* COULOMETRIC ANALYSIS; OVERVOLTAGE. [C.F.M.]

Bibliography: J. J. Lingane, *Electroanalytical Chemistry*, 2d ed., 1958; P. Delahay, *New Instrumental Methods in Electrochemistry*, 1954.

Electrodiagnosis

Three major kinds of electrodiagnostic procedures are used clinically. Electrical activity of the brain is observed and measured by electroencephalography, of the heart by electrocardiography, and of skeletal muscle by electromyography. All three are also used extensively in basic physiological, psychophysiological, and pharmacological research.

Electroencephalography is performed by attaching electrodes to the scalp and, after amplification, recording the detected potential variations (brain wave

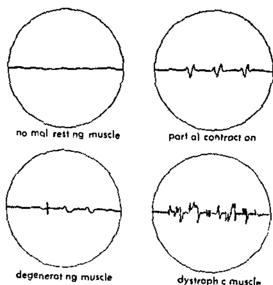
per

foci

raph

means of electrodes contacting the extremities and various points on the thorax wall it is possible to record voltage changes due to heart muscle contraction and so to detect and locate regions within the heart itself where damage has occurred. Some lesions escape notice, but occasionally the technique detects damage when other diagnostic signs are absent. It is useful in prognosis and in following recovery.

Electromyography is the least familiar and newest of the electrodiagnostic procedures mentioned. The objective is to record electrical activity or action potentials of individual motor units and muscle cells at rest and during activity. The motor unit consists of a motor nerve cell in the spinal cord (called an anterior horn cell because of its position), its efferent axon, and in limb muscles the 100 or more muscle cells which it innervates. The detector is a small, usually concentric (coaxial) pair of needle electrodes thrust into the muscle. Electrical activity detected there is amplified fed into an oscilloscope for transient or photographic recording of the e eforms to an inkwriter for direct



Record of synchronous and asynchronous muscle d s- charge

ing or to a loudspeaker for auditory monitoring

Normally a nerve impulse travels from an anterior horn cell out along its axon and terminal axon twigs to excite synchronously the muscle fibers innervated by these twigs The muscle fibers are in turn traversed by their action potentials which cause them to contract It is these synchronized m iscle action potentials which are detected in electromyography

Normally skeletal muscle is electrically silent at rest No oscilloscopic or audible activity is observed Voluntary contraction produces characteristic waveforms and sounds which can be recognized as normal by their amplitude frequency and duration Upon relaxation this activity lessens and disappears (see illustration)

However when the nerve to a m iscle is damaged or interrupted as by pressure stretch or transsection so called fibrillation potentials caused by de synchronized muscular activity appear These are aroused by insertion of the electrode and also occur spontaneously They too can be recognized and are of value in detecting minimal nerve injury and in testing for damage of the nerve supply to deep otherwise clinically inaccessil le muscles Tl e appearance of fibrillation poten ...

... ysteria and in differentiating hysterical paralysis from that caused by real motor nerve damage Finally one of the most important applications of electromyography is in detecting the earliest signs of motor nerve recovery

An older technique corollary to electromyography involves stimulation of nerves or m iscles in th... ove... per... ...y observing re...

sponses to cathodal or anodal stimulation to ma n tained current or repetitive shocks it is poss ble to determine the status of innervation of the muscle

Electrical recording of eyeball movements during sleep promises to give greater insight into the physiology of dreaming and may become useful in psychiatry See BIOPOTENTIALS AND ELECTROPHYSIOLOGY ELECTROCARDIOGRAPHY ELECTROENCEPHALOGRAPHY [J M T]

Bibliography W Dement and N Kleitman ...

9 673–690 1957 R Kovacs Electrotherapy and Light Therapy 6th ed 1949 A A Marinacci Clinical Electromyography 1955 A L Watkins An evaluation of electrodiagnostic testing New Engl J Med 259(18) 868–873 1958

Electrodynamic instrument

A term used in preference to the older term electrodynamometer to refer to an instrument in which the electromagnetic reaction between two co ls or sets of coils each carrying electric current can be measured If the same current flows through all coils in series as in the early Siemens electrodynamometer shown in Fig 1 the instrument can

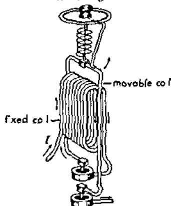


Fig 1 Siemens electrodynamometer (Weston Instruments D v s on of Daystrom Inc)

be calibrated as an ammeter (see AMMETER) Because the reaction is proport onal to the product of the current in the fixed and movable coils the arrangement has found wide usage It can be u ed to indicate ac or dc current voltage or power with a crossed coil movement it can be used for power factor phase angle frequency and capacity measurements For the crossed coil instrument see PHASE METER

Basic instrument A modern version of an electrodynamic instrument mechanism is shown in Fig 2 Note the pair of fixed coils and the movable coil on a vertical staff with spiral countertorque springs for both control and current carrying connections to the moving coil

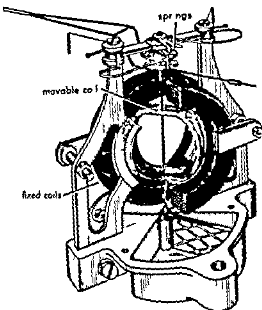


Fig 2 Electrodynamic mechanism (Weston Instruments Division of Daystrom Inc)

The torque developed in an electrodynamic instrument is

$$T = K I_1 I_2 \frac{dM}{d\theta}$$

where I_1 and I_2 are the currents in the fixed and movable coils respectively M is the mutual inductance between the fixed and the movable coils so that $dM/d\theta$ represents the change in mutual inductance or effectively the change in flux coupling as the moving coil rotates K is a constant for the particular system under consideration it includes the number of turns in each set of coils and the electrical geometry of the system

The instantaneous torque is proportional to the instantaneous current in the fixed coils and in the movable coil If the same current flows through both sets of coils the torque is proportional to the square of the current On alternating current the torque is pulsating but the inertia of the moving system causes the pointer to take a position corresponding to the mean value of the torque Thus the deflection is a function of the root mean square (rms) value of the current

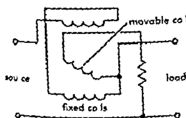


Fig 3 The electrodynamic instrument connected as a wattmeter (Weston Instruments Division of Daystrom Inc)

Transfer instruments Electrodynamic instruments respond equally well to direct or to low frequency alternating currents and therefore are used as transfer instruments to relate dc and ac currents The fundamental standards of the ampere and the volt are direct current standards Transfer instruments are calibrated on direct current using a standard cell and a standard ohm for reference and are then used to measure correctly the rms value of an alternating current

Wattmeters The pointer deflection of the electrodynamic instrument is a function of the product of two currents thus the instrument can be connected as a wattmeter (Fig 3) The main current flows through the fixed coils and a small current proportional to the voltage flows through the movable coil On alternating current where the current may lag the voltage the instrument responds only to the in phase product of the two currents and hence indicates in terms of true watts irrespective of power factor *See* WATTMETER

A particularly refined form of electrodynamic wattmeter is used at the National Bureau of Standards Calibrated on direct current it is used as the standard wattmeter from which all subsidiary standard wattmeters and watthourmeters ac and dc derive their calibration [J H MI]

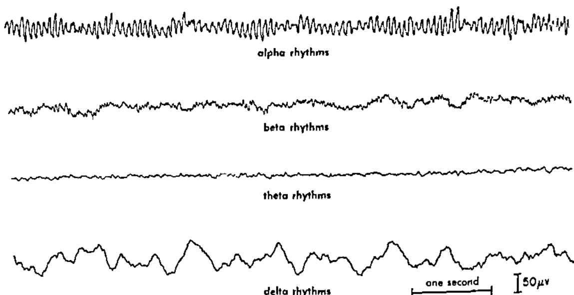
Bibliography F K Harris *Electrical Measurements* 1952 F A Laws *Electrical Measurements* 2d ed 1938

Electrodynamics

The study of the relations between electrical magnetic and mechanical phenomena This includes considerations of the magnetic fields produced by currents the electromotive forces induced by changing magnetic fields the forces on currents in magnetic fields the propagation of electromagnetic waves and the behavior of charged particles in electric and magnetic fields Classical electrodynamics deals with fields and charged particles in the manner first systematically described by James Clerk Maxwell whereas quantum electrodynamics applies the principles of quantum mechanics to electrical and magnetic phenomena Relativistic electrodynamics is concerned with the behavior of charged particles and fields when the velocities of the particles approach that of light Cosmic electrodynamics is concerned with electromagnetic phenomena occurring on celestial bodies and in space *See* COSMIC ELECTRODYNAMICS ELECTROMAGNETISM ELECTRON MOTION IN VACUUM MAXWELL'S EQUATIONS QUANTUM ELECTRODYNAMICS RADIO WAVE PROPAGATION RELATIVISTIC ELECTRODYNAMICS [J W ST]

Electroencephalography

A technique which deals with records of the electrical activity of the brain and their interpretation In most recordings from human subjects the electrical potentials of the brain are seen in an attenuated form because the recordings are made from the scalp through the skull rather t rect



Sample electroencephalograms taken from scalp leads on human subjects

from the brain Direct recordings are secured primarily in experimental animals

The physiological basis of the electrical activity of the brain seems to be potentials associated with the excitatory state of the dendrites The latter are the extensions of the neurons which receive impulses from other neurons and which carry impulses to the cell body from which the axon arises The axon is that part of the nerve which transmits impulses to other nerves or to effector organs *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY

Electroencephalograms Electroencephalograms are records of the brain's electrical activity (see illustration) The potentials recorded are small being of the order of millionths or thousandths of 1 volt Attention has been focused primarily on changing potentials these fluctuate in most parts of the brain from 1 every 2 3 sec to 40/sec or more although in the cerebellum frequencies of 150–250/sec occur Steady potentials have received relatively little study

Electrical activity of the brain first appears in the developing organism at the time when clusters of nerve cells called nuclei can be distinguished visually from one another and shortly after that small particles called Nissl granules are detectable within the individual nerve cells Once the activity appears it generally becomes progressively faster with increasing age until maturity is reached Some rhythms decrease in frequency or appear and then disappear at a later age

Alpha rhythms Hans Berger from whose discoveries present day electroencephalography stems named the first recurrent waves he detected alpha rhythms This name is given at present to rhythms which have a pulse frequency of 8–13/sec and which are localized primarily but not exclusively in the parieto occipital region They are customarily found in the normal human adult when he is relaxed and has his eyes closed

Beta rhythms Low voltage fast activity in the range of 13 30 pulses/sec is called beta rhythm Beta rhythm is often encountered when a person is aroused and anxious

Theta rhythms Rhythms which recur at the rate of 4–7/sec are known as theta waves because originally they were thought to originate in the thalamus They too, have been associated with affective states and are often found in adolescents with behavior disorders Despite the name the site of origin of these waves is probably the hippocampus

Delta rhythms Delta rhythms slow waves with frequencies of 0 5–3/sec appear in the normal subject when he is asleep They have appreciably greater amplitude than do the waves referred to above and their focus is a forward position in the brain Prior to the subject's falling asleep 14 spindles per second are often recorded from the central region of the brain These occur in groups lasting only a few seconds During sleep low voltage fast activity may appear and if this is accompanied by rapid movements of the eyes the subject is probably dreaming If awakened at this time subjects will report dreaming 85% of the time Dream periods occur about five times each night becoming progressively longer as sleep continues

Relationship to learning If alpha rhythms are being recorded from a human subject and a novel stimulus is presented to him the rhythms will cease In recordings made from electrodes placed directly in the brain instead of on the scalp the disappearance of the alpha rhythms is accompanied by a shift to low fast random activity which can be seen in various areas of the brain

This shift constitutes an alerting response If the stimulus is repeated several times the response commonly disappears gradually Some stimuli such as an electric shock regularly evoke an alerting response the brain adapts to them slowly if

at all If a stimulus of the shock type is paired with one to which an animal has ceased to respond the original response will be reinstated and will continue to be evoked This happens not only in the waking state but also during sleep The awakening of a mother by her child's cry but not by other voices may be mediated by this process Similarly attention may be a function of this mechanism

If an animal has been trained not to respond to a stimulus the appearance of the stimulus is accompanied by sleep spindles in the motor cortex This finding appears to support Pavlov's ideas of inhibition and sleep There is also evidence that in the course of learning areas of the brain originally unresponsive to a stimulus may come to respond to it Other evidence indicates that the conditioned [illegible]

cur within the brain but even in the absence of seizures peculiar waves are noticeable in the electroencephalogram of epileptics During seizures various forms of epilepsy can be distinguished According to F A Gibbs and coworkers grand mal epilepsy is characterized by extreme acceleration of the electrical activity of the cortex psychomotor attacks by extreme slowing of this activity and petit mal by alternation of fast and slow activity Electroencephalography is also useful in the diagnosis of epilepsy and in locating the site of the abnormality

Electroencephalography is used in the detection of cerebral tumors Slow fairly rhythmic electrical potentials occur in the regions of the scalp overlying the affected area These waves are the same as those [illegible] gene[illegible] the [illegible]

BRAIN LEARNING THEORIES NEUROPHYSIOLOGY

[J P F]

Bibliography F A Gibbs E L Gibbs and W G Lennox Cerebral dysrhythmias of epilepsy *AMA Arch Neurol Psychiat* 39 298–314 1938

Electrojet, upper air

An intense electric current assumed to occur occasionally in restricted areas of the atmosphere at altitudes of about 100 km or higher In some theories which are devised to account for variable features of geomagnetism and earth currents (particularly electromagnetic storms) electrojets are postulated as the immediate source of some aspects of the variations The regions in which the electrojets occur are supposed to have a much greater electrical conductivity than other regions at the same altitude This admits of an intense electric flux if an adequate electromotive force is at hand The electromotive force is probably generated by winds which move the conducting air of extensive surrounding areas across the earth s permanent magnetic field *See* GEOMAGNETISM STORM

[O H G]

Electrokinetic phenomena

Phenomena associated with the movement of charged particles through a continuous medium or with the movement of a continuous medium over a charged surface The four principal electrokinetic phenomena are electrophoresis electroosmosis streaming potential and sedimentation potential or Dorn effect These phenomena are related to one another through the zeta potential ζ of the electrical double layer which exists in the neighborhood of the charged surface

Electrically charged layers The distribution of electrolyte ions in the neighborhood of a negatively charged surface and the variation of potential ψ with distance from the surface are shown in Fig 1 According to O Stern two different layers of ions are associated with the charged surface The layer of ions immediately adjacent to the surface is called the Stern layer The ions of this layer are held to the charged surface by a combination of electrostatic attraction and specific adsorption forces such as short range van der Waals interactions and chemical bonds The thickness δ of this layer is assumed to be equal to the ionic radius of the adsorbed ion species The second layer of ions is the Gouy layer The boundary between the two layers is the limiting Gouy plane The ions in the Gouy layer are acted upon only by electrostatic forces and thermal motions of the liquid environment (Brownian motion) and they form a diffuse atmosphere of opposite charge (positive charge in Fig 1) to the net charge at the limiting Gouy plane The net charge density of the diffuse ion

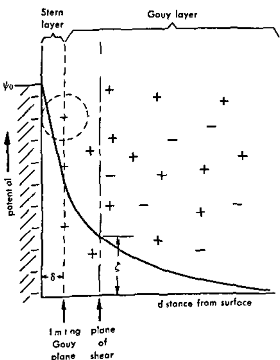


Fig 1 Electrical double layer

atmosphere of the Gouy layer decreases exponentially with distance from the limiting Gouy plane The Gouy layer forms the positive half of an electrical double layer and the charged surface plus the Stern layer form the negative half The effective distance of separation $1/\kappa$ between the two halves of the double layer is determined by the concentration of electrolyte (ionic strength) For an electrolyte of univalent ions in water at 25°C the relationship for $1/\kappa$ from the Debye Huckel theory is

$$\frac{1}{\kappa} = \frac{3 \times 10^{-8}}{\sqrt{c}} \qquad (1)$$

where c is the concentration of electrolyte (moles/liter)

The variation of potential ψ with distance x from the charged surface is shown by the solid curve in Fig 1 ψ_0 represents the thermodynamic reversible electrode potential which is independent of the properties of the electrical double layer and dependent only on the activity of the ion which is in reversible electrochemical equilibrium with the substance of the charged surface The potential ψ decreases linearly with increasing distance x in the region of the Stern layer In the region of the Gouy layer ψ decreases exponentially with increasing distance x as shown by G Gouy and W Chapman

Displacement of charged layers In the four listed electrokinetic phenomena a displacement occurs at some plane (plane of shear) between the charged surface and its atmosphere of ions The position of the slipping plane in Fig 1 is shown to be located in the Gouy layer The potential at the plane of shear is the ζ potential From the theories of Gouy and Chapman for spherical particles

$$\zeta = \frac{q}{Da}\left(\frac{1}{1+\kappa a}\right) \qquad (2)$$

where $1/\kappa$ is the effective thickness of the double layer q the net charge of the particle inside the plane of shear D the dielectric constant of the liquid and a the particle radius at the plane of shear For flat surfaces

$$\zeta = \frac{4\pi e}{D\kappa} \qquad (3)$$

where e is the charge per unit area of surface Equations (2) and (3) show that ζ potential is determined by the net charge at the plane of shear and $1/\kappa$ the effective thickness of the ion atmosphere In turn ζ potential controls the rate of transport between the charged surface and the adjacent liquid The relationship between rate of transport v_E and ζ potential which is valid for all four electrokinetic phenomena is

$$v_E = \frac{D\zeta E}{4\pi\eta} \qquad (4)$$

where v_E is the velocity of the liquid at a large

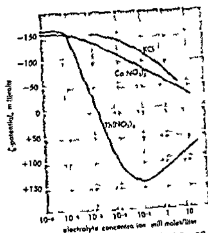


Fig 2 Effect of electrolyte concentration on ζ potential

distance from the charged surface E the field strength (volts/cm), and η is the viscosity of the liquid The conditions for validity of Eq (4) must be that the double layer thickness ($1/\kappa$) must be small compared to the radius of curvature of the surface the substance of the surface must be nonconducting and the surface conductance of the interface must be negligible The equations which relate ζ potential to electroosmotic flow rate and streaming potential may be obtained from Eq (4) by use of Poiseuille's law for laminar flow through a capillary For electrophoresis and sedimentation potential (Dorn effect), v_E is the velocity of the particles E is the applied field strength for elec

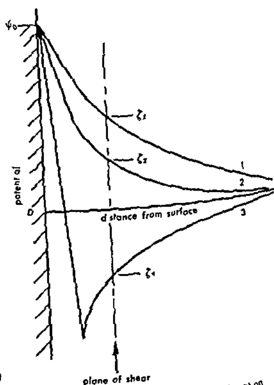


Fig 3 Reversal of ζ potential by ion adsorption

trophoresis whereas it is the gradient of potential developed by the sedimentation of charged particles in the Dorn effect

Electrophoresis electroosmosis and streaming potential experiments have been shown to yield identical ζ potentials for several different interfaces particularly glass-water and protein water systems The sedimentation potential has not been significantly studied

The effect of electrolytes on the ζ potential of glass-water interfaces is shown in Figs 2 and 3 As shown in Fig 2 an increase in electrolyte concentration produces a decrease in ζ potential and ions of high charge of opposite sign to that of the surface can completely reverse the sign of the ζ potential The explanations for these two effects are given in Fig 3 where the variation in ψ with distance from the surface is shown for low concentration of electrolyte in curve 1, moderate concentration of electrolyte in curve 2, and charge reversal by adsorption of ions (Th^{4+} on glass) in curve 3 Curves 1 and 2 show that an increase in electrolyte concentration reduces ζ potential by reducing $1/\kappa$ as indicated by Eqs (1), (2) and (3) Curve 3 shows that reversal of charge by ion adsorpt [illegible]

gives [illegible]

origin [illegible]

ELECTROPHORESIS, STREAMING POTENTIAL

[Q V W]

Electrokinetic transducer

An instrument used to convert dynamic physical forces such as vibration and sound to electric power the instantaneous values of which have scaled correspondence to the instantaneous forces

The electrokinetic phenomenon also known as the streaming potential occurs when a polar fluid is moved through an insulated capillary The electrokinetic transducer uses a closed volume of a polar fluid with a permeable refractory ceramic or fritted glass member making two chambers usually of equal volumes Electrodes are placed on each side of this member and extend outside each of the chambers Electric potential changes caused by the streaming potential are measured at these electrodes See ELECTROKINETIC PHENOMENA, STREAMING POTENTIAL

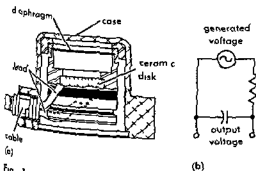


Fig 1 Electrokinetic transducer (a) Cross-sectional view (b) Simplified electrical equivalent circuit

Most electrokinetic transducers are constructed by placing a case around a *unit cell* (Fig 1*a*) This unit cell is constructed in the same manner for use in microphones accelerometers and dynamic pressure pickups When the transduction principle [illegible]

electrokinetic transducer is shown in Fig 1*b*

A dynamic pressure pickup may have a range of 0 01–1000 psi A microphone may have a range of 70–216 db The low end of the range is limited by the inherent noise characteristics caused by intrinsic fluid movement

The electr [illegible]

Linear accelerometer [illegible] 0 500 [illegible]

stant peak magnitude are applied to a unit cell transducer Fig 2 describes a typical response curve

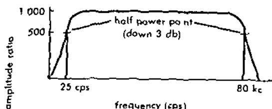


Fig 2 Frequency response curve of typical electrokinetic transducer

The expression derived by Helmholtz in 1879 shows the relationship between streaming potential and an applied pressure

$$\frac{E}{P} = \frac{\zeta\epsilon}{4\pi\nu\sigma}$$

where ζ is zeta potential ϵ is dielectric constant ν is coefficient of viscosity σ is fluid conductivity, E is streaming potential and P is applied pressure

From this expression one can see that for a pressure gage the best fluid must have (1) high zeta potential (or polar moment) (2) high dielectric constant (3) low coefficient of viscosity, and (4) low electrical conductivity

Other transducers follow similar formulas The zeta potential is dependent on the materials in contact with the fluid and on the polar moment of the fluid

The polar fluids are rated in order of their available power temperature span between boiling and freezing points toxicity resistivity and reactions with other materials Acetonitrile propionitrile and nitrobenzene are among the fluids used These fluids are compatible with materials such as soft glass alumina, corrosion resistant steels and aluminum

Other materials react unfavorably with the polar fluids and degrade the performance [D N M]

Bibliography S Glasstone *Introduction to Electrochemistry* 1942 R A Gortner Jr and W A Gortner *Outline of Biochemistry* 3d ed 1949 R J W LeFevre *Dipole Moments* 3d ed 1953

Electroluminescence

The production of light emission in a suitable solid phosphor by the application of an electric field Electroluminescence often called the Destriau effect after its discoverer G Destriau differs from other common means of light production in that electrical energy is converted into light without an intermediate stage Usually the electric field is an alternating one and the light also fluctuates

Electroluminescent lamps The incandescent lamp uses electrical energy to heat the filament and the emitted light is a result of the heat inevitably much of the energy radiated by the lamp is outside the useful range of the visible spectrum The fluorescent lamp also involves an intermediate step for the electrical energy first creates a discharge in the mercury vapor inside the lamp The gaseous discharge emits ultraviolet light (principally at 2537 A) this light excites a phosphor coated on the inside of the glass tube and the phosphor then emits light in the visible region Although the fluorescent lamp is complex it is efficient because it does not waste very much energy producing heat The white fluorescent lamp emits 70 lumens per watt (lu/watt) which is close to its maximum theoretical efficiency while the incandescent lamp [illegible] g about 5 lu/watt however they have a theoretical upper limit of more than 100 lu/watt These lamps are nevertheless useful in low light level applications and in cases where a large area lamp has advantages It seems likely that as the phenomenon of electroluminescence becomes better understood and engineering techniques improve the efficiency of the lamp may be increased

To make a solid phosphor such as zinc sulfide (ZnS) powder electroluminesce efficiently it is necessary to prepare it so that a thin layer of a semiconductor such as cuprous sulfide (Cu_2S) forms on the powder It is believed that a thin highly insulating layer is formed at the contact between the ZnS and the Cu_2S Most of the voltage drop from the applied electric field occurs across this layer and the field strength in the layer may approach that necessary for electrical breakdown Free electrons are thus accelerated so that they excite luminescent centers or produce additional free electrons which eventually give up their energy to luminescent centers (A luminescent center is a region usually near imperfections or chemical impurities where electrons can congregate and have an arrangement unlike that in the lattice as a whole *See* LUMINESCENCE) An analysis of this situation shows that the luminescent intensity I should vary with the voltage V, as

$$I = A \exp(-B/V^{1/2})$$

where A and B are constants This has been verified over a very wide range of brightness For a diagram of an electroluminescent lamp and additional details *see* LIGHT PANEL

Light amplifiers From the preceding equation it may be seen that a large change in electroluminescent brightness may occur with a relatively small change in applied voltage This property has led to the development of light amplifiers using electroluminescence

If radiation falls in a pattern on an electroluminescent panel with no voltage applied and if the radiation is of such a wavelength that free electrons are produced a normal luminescent image of the radiation pattern will appear on the panel If an increasing ac voltage is then applied to the panel the image will grow in intensity until it is many times the original brightness In this case the electroluminescence starts with electrons made free by electromagnetic radiation rather than depending on thermally freed electrons The device thus not only can convert light from one frequency to another (for example ultraviolet to visible light) as any phosphor will do but also can emit more radiant energy than is incident on it which a normal phosphor cannot do In addition to being a light amplifier the device has been used with x rays since they also produce free electrons. In the usual medical fluoroscopic examination the x rays that penetrate the patient fall on a fluorescent screen (*see* RADIOLOGY) To minimize x ray exposure to the patient low x ray intensities are used and the radiologist adapts his eyes to dark to see the low brightness image If an electric field is applied to a specially designed fluoroscopic screen, a brighter image results allowing fluoroscopic examinations to be made more conveniently and with a reduction in x ray exposure of the patient Similarly electroluminescent panels can be made to respond to infrared radiation For details on another type of light amplifier *see* LIGHT AMPLIFIER

Carrier injection luminescence An effect due to an electric field may also be considered under the subject of electroluminescence although it is less commonly encountered It sometimes occurs when for instance a thin wire touches a semiconductor If the semiconductor is *p* type that is if it has free holes (places where electrons are missing) and if the voltages are arranged so that electrons from the wire are injected into the semiconductor the electrons and holes may recombine to give off light This process often called *carrier injection luminescence* is observed in silicon carbide which emits light in the visible spectrum and in other semiconductors which usually emit in the infrared It is normally a very inefficient light production process since most of the electrons and

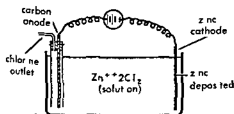
holes recombine in ways which do not involve luminescence

Gudden Pohl effect Another effect due to an electric field is called the Gudden Pohl effect after its discoverers B Gudden and R Pohl If an electric field is applied to a phosphorescent material as its light is slowly decaying the electrons in traps" (defects or chemical impurities which capture electrons) are pulled out more rapidly and used to produce luminescence In this way the phosphorescence decay is accelerated giving an increase in light intensity and a decrease in decay time Light emission ceases when all the electrons are pulled out of their traps [C C K J H S]

Bibliography D Curie *Progress in Semiconductors Theories of Electroluminescence* vol 2 1957 G Destriau *Trans Faraday Soc* 35 227 1939 W W Piper and F E Williams *Solid State Physics Electroluminescence* vol 6 1958

Electrolysis

A method by which reactions are carried out in solutions of electrolytes or in molten salts by use of electricity The electrodes of an electrolytic cell are immersed in an electrolyte solution or in molten salts and are connected to a direct current power supply One or several reactions occur at each electrode when current flows through the cell Reduction a reaction in which electrons are consumed occurs at the electrode called the cathode oxidation occurs at the anode For instance sodium is produced at the cathode by reduction and chlorine at the anode by oxidation in the electrolysis of molten sodium chloride



Electrolysis of zinc chloride solution

Applications are important and varied industrial production of chemicals metallurgical extraction of metals electroplating of metals metal finishing and production of electricity in batteries Metallic corrosion often involves electrolytic processes For application to analytical chemistry *see* ELECTRODEPOSITION ANALYSIS POLAROGRAPHIC ANALYSIS

Theory The quantity of electrolysis products their rate of production and quite often their nature depend on electrolysis conditions According to Faraday s law the quantity of substance being consumed or produced by a single electrode reaction is proportional to the quantity of electricity consumed in electrolysis This quantity of electricity is equal to the product of the current multiplied by the duration of electrolysis for a constant current or to the integral of the current over the duration of electrolysis for a variable current *See* COULOMETER ELECTROCHEMICAL EQUIVALENT

The nature and the relative abundance of electrolysis products at each electrode generally depend on the electrode potential *See* ELECTROMOTIVE FORCE (CELLS) Direct control of potential is rarely used and electrode potentials in industrial cells are controlled indirectly by adjustment of the current density (current per unit area) at each electrode Control is achieved because the current density depends on the electrode potential (*see* DECOMPOSITION POTENTIAL OVERVOLTAGE) Control of the electrolysis current has the advantage of allowing connection of several identical electrolytic cells in series in industrial installations

Electrolytic cells are characterized by their current efficiency and their power consumption efficiency The current efficiency is the ratio of the quantity of a substance being consumed or produced to the theoretical quantity of this substance as calculated from Faraday s law The current efficiency of a single electrode reaction occurring without losses such as side reaction electrolysis of the solvent evaporation and so forth is 100%

The second efficiency characteristic of an electrolytic cell the power consumption efficiency is the ratio of the theoretical electrical power to the actual electrical power that is consumed in the production or consumption of a given quantity of substance The power consumption efficiency is smaller than 100% because of overvoltage phenomena losses of products by side reactions and ohmic drop (voltage drop) in the cell Power efficiencies as low as 50% or even lower are not uncommon

Applications Industrial applications for inor

trolysis Water is enriched in deuterium oxide (heavy water) by electrolysis (There is isotopic separation because the overvoltage for discharge of deuterium ions is larger than for hydrogen ions) Certain metals such as aluminum magnesium and sodium are produced by electrolysis of molten salts Deposition of these metals from aqueous solution is impossible because this reaction requires higher cathodic potentials than hydrogen evolution Likewise fluorine is produced by oxidation of fluoride ion in anhydrous hydrofluoric acid electrolysis of aqueous solutions of fluoride produces oxygen because this reaction occurs at lower anodic potentials than fluorine evolution

Electroplating of thin layers of a corrosion resistant metal on fabricated objects is an important technique Chromium and nickel are most commonly used but electroplating of other metals such as gold silver and copper also has applications The composition of the electrolytic bath influences the structure and surface finish of the metallic coating and numerous formulas involving met

complexes and organic additives have been developed empirically Electroplating is also applied to the industrial refining of copper silver gold nickel Certain metals (copper zinc cadmium) are extracted from low grade ores by electrolysis of a solution of their ores (electrowinning) The opposite reaction of electroplating—anodic oxidation—is applied to metal finishing in electropolishing

Electrolysis of organic compounds has found only few industrial applications although numerous electrode reactions have been studied Purely chemical preparative methods are more economical and often simpler than electrolysis In some cases however electrolysis involves reactions which are more easily controlled than purely chemical methods Important reactions include reduction of nitro compounds aldehydes ketones carboxylic acids unsaturated compounds halogenated substances oxidation of fatty acids (Kolbe reaction) alcohols aldehydes ketones sugars and halogenation by anodic oxidation *See* ELECTROCHEMISTRY ELECTROMETALLURGY ELECTROPLATING OF METALS

[P D]

Bibliography M J Allen *Organic Electrode Processes* 1958 A G Gray (ed) *Modern Electroplating* 1953 C L Mantell *Industrial Electrochemistry* 3d ed 1950

Electrolyte

A chemical compound which when fused or dissolved in certain solvents usually water will conduct an electric current The passage of the current is always accompanied by decomposition of the electrolyte called electrolysis which takes place at the electrodes All electrolytes in the fused state or in solution give rise to ions which conduct the electric current The phenomena of electrolysis are summarized by Faraday s laws of electrolysis All acids bases and salts are electrolytes *See* CURRENT ELECTRIC ELECTRODE ELECTROLYSIS ION SOLVENT

Electrolytes are divided into strong and weak electrolytes Strong electrolytes usually contain a stable ionic bond and are wholly ionized in solution and usually in the crystalline state Weak electrolytes are *only partially* ionized in solution Metallic hydroxides and salts are usually strong electrolytes for example potassium hydroxide and sodium chloride A weak electrolyte contains a covalent bond which on dissolving in a solvent such as water may be transformed *into* an ionic bond A solution of a weak electrolyte contains both the ionic and the covalent forms in equilibrium for example acetic acid in water consists of a mixture of undissociated molecules CH_3COOH and of the ions CH_3COO^- and H^+ *See* CHEMICAL BINDING EQUILIBRIUM IONIC

[T C W]

Electrolytic conductance

The transport of electric charges under electric potential differences by particles of atomic or larger size This phenomenon is distinguished from metallic conductance which is due to the movement of electrons The charged particles that carry the electricity are called ions

Positively charged ions are termed cations sodium ion Na^+ is an example The negatively charged chloride ion Cl^-, is typical of anions The negative charges are identical with those of electrons or integral multiples thereof The unit positive charges have the same magnitude as those of electrons but are of opposite sign Colloidal particles which may have relatively large weights, may be ions and may carry many positive or negative charges Electrolytic conductors may be solids, liquids or gases There are a few conductors called semiconductors, with properties that are intermediate between the metallic and electrolytic types

Measurement Conductances are usually reported as specific conductances κ which are the reciprocals of the resistances of cubes of the materials 1 centimeter (cm) in each dimension placed between electrodes 1 cm square on opposite sides These units are sometimes called mhos that is ohms spelled backward Conductances of solutions are usually measured by Friedrich Kohlrausch's method in which a Wheatstone bridge is employed Such a bridge is shown diagrammatically in Fig 1 The resistances R_3 and R_4 (usually of the same value) form two arms of the bridge as shown Resistance R_2 is adjustable and the remaining arm is the cell holding the electrolytic conductor or as is usually stated solution of electrolyte Direct current and the usual galvanometers cannot be used since the passage of current produces chemical reactions which cause polarization at the electrodes of the cell thus modifying the solution and its conductance However by using a source of alternating current the polarization may be avoided The electrochemical reactions occurring when the current is briefly passed in one direction may be reversed when the direction of the current is changed The electrodes of the cells are almost universally made of platinum and are platinized that is given a coating of finely divided platinum plated out from a solution of platinic chloride This greatly increases the active surface of the electrode and tends to reduce polarization Electronic oscillators capable of yielding alternating current of definite frequencies have been used

To determine the conductance C that is the reciprocal resistance $1/R$ of the cell of Fig 1 the resistance R_2 of the bridge is adjusted until a

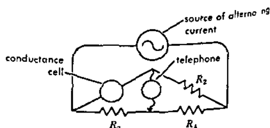


Fig 1 Wheatstone bridge circuit for measuring electrolytic conductance R_3 and R_4 are fixed resistances R_2 is a variable resistance

Fig 2 Conductance cell

minimum of sound is heard in the telephone. Greater sensitivity may be obtained by inserting an electronic amplifier between the telephone and the bridge. When the bridge has been adjusted to give the minimum sound, the conductance is given by the relation $C = R_4/R_2R_3$. From this the specific conductance κ may be obtained from the equation $\kappa = KC$ in which K is the cell constant. Occasionally this constant can be computed from the dimensions of the cell. Usually, however, it is determined by using a solution whose κ value is accurately known from measurements in such a cell or, as was done by G. Jones and B. C. Bradshaw, by comparison with the specific conductance of mercury.

For precision work, care must be taken to avoid errors due to electrical reactances. This has been done in bridges designed by Jones and R. C. Josephs and by Theodore Shedlovsky. A typical properly designed conductance cell is shown in Fig. 2. The cell is filled with solution through the center tubes. Electrical contact is made with the electrodes by platinum wires sealed through the glass wall. These connect the mercury in the outside tubes, which are widely spread as shown to avoid errors due to electrical capacity.

The question has been raised as to whether Kohlrausch's method for determining conductances, involving polarizable electrodes and high frequency currents, gives the same values as would be obtained by a method using nonpolarizable electrodes and direct current. Careful experiments to test this matter ha[illegible] urec[illegible] and[illegible] meth[illegible]

Equivalent conductance Although many substances and mixtures show electrolytic conductance, the greater part of the research on the subject has been on aqueous solutions of salts, acids, and bases. There has been considerable data accumulated for solutions of such electrolytes in nonaqueous solvents, such as alcohols. The data are usually given in terms of equivalent conductance Λ, which is defined by the expression

$$\Lambda = \frac{1000\kappa}{c}$$

in which κ is the specific conductance and c is the concentration in equivalents per liter. Values of Λ change with the concentration and in general increase as the solutions measured are made more dilute, that is, as c is decreased. A plot of values of the equivalent conductance Λ against $\sqrt{c}$ for some typical electrolytes is shown in Fig. 3. Svante Arrhenius, who was the first to assume that electrolytic conductance is due to freely moving charged ions, explained the decrease of Λ with increasing c by assuming that the number of ionic carriers gets smaller as the concentration increases, and he computed a degree of dissociation α by the formula

$$\alpha = \frac{\Lambda}{\Lambda_0} \tag{1}$$

The term Λ_0 is obtained by determining Λ at a series of low concentrations and extrapolating to a limiting value, termed the equivalent conductance at infinite dilution. Though Eq. (1) has been shown by later work to give nearly the right values of α for certain poorly conducting solutions, it is now considered to be much in error for the so-called strong electrolytes. These include most salts, such as potassium chloride, KCl, and sodium sulfate, Na_2SO_4, and inorganic acids and bases, such as hydrochloric acid, HCl, and sodium hydroxide, NaOH. For an electrolyte which yields two types of ion, it can be shown that

$$\Lambda = F\alpha(U^+ + U^-) \tag{2}$$

in which F is the faraday and U^+ and U^- are the mobilities or speeds under unit potential difference of the positive and negative ions, respectively. For Eq. (1) to hold, these mobilities must be constant from the equivalent concentration c at which Λ is measured to infinite dilution. This requires that the transference numbers of the ions be constant in the same range, which is seldom the case.

Fig 3 Determination of equivalent conductance infinite dilution

Since the advent of the Debye Huckel theory of interionic attractions strong electrolytes have been considered to be completely dissociated that is the term α of Eq (2) is equal to unity for these substances The decreases observed in the values of the equivalent conductances Λ with increases in concentration are assumed to be due to reductions in the values of the ionic mobilities U and U According to the theory of P Debye and E Huckel the ion possesses an ionic atmosphere distributed with radial symmetry around the ion as center This is due to the fact that interionic attractions and repulsions together with thermal vibrations tend to produce a slight preponderence of negative ions around a positive ion and vice versa The presence of this atmosphere leads to two effects both of which result in the lowering of ionic mobilities with increasing ion concentrations These are the electrophoretic effect and the time of relaxation effect The first of these is presumed to produce a motion of the solvent opposite to that of the ion The time of relaxation effect may be briefly outlined as follows Around a selected ion the ionic atmosphere has spherical symmetry However if the ion is moved the ionic atmosphere will move with it but adjustment is not quite instantaneous and the unadjusted portions of the atmosphere produce a braking action on the central ion The adaptation of the Debye Huckel theory for conducting solutions is due to Lars Onsager His equation for very dilute uni univalent electrolytes such as sodium chloride (NaCl) is

$$\Lambda = \Lambda_0 - (\theta\Lambda_0 + \sigma)\sqrt{c} \qquad (3)$$

in which θ and σ are given for uni univalent electrolytes by

$$\theta = \frac{8\ 16 \times 10^5}{(DT)^{1\ 2}} \quad \text{and} \quad \sigma = \frac{8\ 28}{\eta(DT)^{1\ 2}}$$

in which D is the dielectric constant at the absolute temperature T and η is the viscosity

Equation (3) yields accurate values of the data for strong electrolytes up to concentrations of about 0 001 molar above which there are small deviations Modifications of Eq (3) for solutions of salts of higher valence types such as calcium chloride $CaCl$ and lanthanum chloride $LaCl_3$ are available and have also been found to agree with the data for dilute solutions

Onsager in his derivation of Eq (3) treated the ions as point charges Later Raymond Fuoss and Onsager extended the theory to include the radii of the ions and also the effects of higher concentrations The ion sizes obtained from conductance data agree closely with those calculated from activity measurements See ACTIVITY (THERMODYNAMICS)

Equation (3) or its empirical and theoretical ext-
the
tior

conductances Λ_0 of typical strong electrolytes in aqueous solution at 25°C are listed below

HCl	426 16	KBr	151 9	NaOH	24 8
LiCl	115 03	NaI	126 94	$MgCl_2$	129 40
NaCl	126 45	KI	150 38	$CaCl_2$	135 84
NH_4Cl	149 86	KNO_3	144 96	Na_2SO_4	129 9

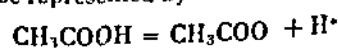
Ion conductances Values of the limiting equivalent conductance Λ_0 may be assigned to each of the ions of an electrolyte Thus for potassium chloride $\Lambda_{0\ KCl} = \lambda_{0\ K^+} + \lambda_{0\ Cl}$ and the value of $\lambda_{0\ Cl}$ is the same whether it is derived from measurements on HCl NaCl or KCl solutions This additive relation is known as Kohlrausch s law of the independent mobility of ions However it is necessary to obtain the value of λ_0 for at least one ion constituent independently in order to establish the ion conductances of the other ions The relation used is

$$t_0^+\Lambda_0 = \lambda_0^+ \quad \text{or} \quad t_0\ \Lambda_0 = \lambda_0$$

in which Λ_0 is the limiting equivalent conductance of an electrolyte and t_0^+ and t_0 are the limiting transference numbers of the positive and negative ion constituents respectively (*see* TRANSFERENCE NUMBER) D A MacInnes L G Longsworth and T Shedlovsky have shown that the same value of $\lambda_{0\ Cl}$ within 0 02% is obtained from precision conductance and transference measurements on solutions of hydrogen lithium sodium and potassium chlorides Values of the limiting ionic conductance at 25°C are given below for some ions

K^+	73 52	Cl^-	76 34
Na^+	50 11	Br^-	78 4
H^+	349 82	I^-	76 8
Ag^+	61 92	NO_3^-	71 44
Li^+	38 69	HCO_3^-	44 48
NH_4^+	73 4	OH^-	198
½Ca^{++}	59 50	$CH_3CO_2^-$	40 9
½Ba^{++}	63 64	$CH_2ClCO_2^-$	39 7
½Mg^{++}	53 06	½SO_4^{--}	79 8

Ionization constants So far in this discussion only solutions of strong electrolytes have been considered There are many solutions of electrolytes particularly of acids and bases but including some salts for which it is necessary to assume incomplete dissociation into ions for example aqueous solutions of acetic acid CH_3COOH The equilibrium between the ions and the undissociated portion may be represented by



$$CH_3COOH = CH_3COO^- + H^+$$

If the law of mass action is applied to this equilibrium the expression

$$K = \frac{c\alpha^2}{(1-\alpha)} \qquad (4)$$

is obtained in which c is the concentration α is the degree of dissociation and K is the ionization constant This is known as Ostwald s dilution law Using Eq (1) Arrhenius relation $\alpha = \Lambda/\Lambda_0$ a

fairly constant value of K is observed from the data of MacInnes and Shedlovsky for a wide range of values of the concentration c as is shown from the data in the table. However, there is a slight but unmistakable increase of K as the concentration is increased.

If a similar computation is made using data on conductivity for a stronger acid, such as chloroacetic acid, a much more rapid increase in K with the concentration is observed, and with the figures for hydrochloric acid a very wide variation of K is observed. In these computations it has been assumed that the mobilities of the ions do not change with the concentration and that the ions are perfect solutes; neither assumption is justified. A better value of the degree of dissociation is obtained by the relation $\alpha_1 = \Lambda/\Lambda_e$, in which Λ_e is the computed equivalent conductance of the completely dissociated acid at the ion concentration $c\alpha_1$. The data for this computation are obtained from measurements on strong electrolytes and involve a series of approximations. Values of K' obtained from Eq. (4) using α_1 are shown in column 4 of the table. To take account of the effect of the ionic atmosphere on the properties of the ions, Eq. (4) must be modified as follows:

$$K'' = \frac{c\alpha_1^2}{(1-\alpha_1)}f^2 \qquad (5)$$

in which f is the activity coefficient. In its simplest form the Debye-Huckel theory yields $f = 0.5056\sqrt{c\alpha_1}$ at 25°C. In the last column of the table values of K'' are listed. It will be seen that they are constant. A corresponding constancy is observed when data for stronger acids, such as chloroacetic, are treated in a similar manner. It is thus seen that the interionic attraction theory is useful in interpreting the data for weak, as well as for strong, electrolytes.

Ionization constant of acetic acid at 25°C.

Concentration, equiv/liter, $c \times 10^3$	Equivalent conductance	Ionization constants $\times 10^5$		
		K	K'	K''
0.021814	210.38	1.760	1.768	1.752
0.15321	112.05	1.767	1.778	1.752
1.02831	48.146	1.781	1.797	1.751
2.41400	32.217	1.789	1.809	1.750
5.91153	20.962	1.798	1.823	1.748

As with strong electrolytes, deviations from the simple theory are observed for more concentrated solutions. These can be partly accounted for by higher terms in the Debye-Huckel theory. However, A. Katchalsky, H. Eisenberg, and S. Lifson have shown that a more important effect is a dimerization or doubling of the molecules of carboxylic acids as the concentration increases.

Nonaqueous systems. In addition to the study of water solutions of electrolytes, considerable study has been given to electrolytes in nonaqueous and mixed solvents. In general, the same principles as those outlined above apply to the interpretation of the results. However, fewer of the electrolytes are completely dissociated, and the degrees of dissociation of the weaker acids and bases are lower. This is due to the fact that in general the dielectric constants of nonaqueous solvents are smaller than those of water, with the result that the attractions between positive and negative ions are greater.

It will be observed that the discussion given above is confined to quite dilute solutions of electrolytes. For concentrated solutions few generalizations of any value can be given. Fused salts have quite large conductances, but here again little in the way of theoretical explanation of the results is yet available.

If, instead of using quite low potentials in the measurement of electrolytic conductances, voltages of the order of 100,000 are employed, the conductances observed are no longer constant but tend to increase with the potential used. Under these conditions Ohm's law evidently is not valid. This increase of conductance with high potentials is called the Wien effect. This effect is in accord with the interionic attraction theory. When the velocity of the ions becomes sufficiently great, the ion atmospheres do not have time to form to their full extent, with the result that both the electrophoretic and time of relaxation effects exert less influence on the conductance. However, a large Wien effect is also found for weak acids and bases. It would appear that the high potentials produce temporarily additional ionization of these substances. This explanation has been proposed and discussed theoretically by Onsager. If very high frequencies are used in the measurements, an increase in the conductance, termed the Debye-Falkenhagen effect, is observed. This can also be explained by the interionic attraction theory. See ELECTROCHEMISTRY; ELECTROMOTIVE FORCE (CELLS). [D.A.M.]

Bibliography: H. S. Harned and B. B. Owen, *Physical Chemistry of Electrolytic Solutions*, 3d ed., 1958; R. A. Robinson and R. H. Stokes, *Electrolyte Solutions*, 1955.

Electrolytic tank

A special type of computing machine which owes its existence to the fact that in ideal fluid flow the velocity potential φ and in planar flow the stream function ψ satisfy Laplace's equations (see FLUID-FLOW PRINCIPLES; FLUID-FLOW PROPERTIES; HYDRODYNAMICS). Laplace's equations in rectangular cartesian coordinates are written

$$\frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} = 0$$
$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = 0 \qquad (1)$$

with velocity components

$$u = \frac{\partial\varphi}{\partial x} = \frac{\partial\psi}{\partial y} \qquad v = \frac{\partial\varphi}{\partial y} = -\frac{\partial\psi}{\partial x} \qquad w = \frac{\partial\varphi}{\partial z} \qquad (2)$$

Because the voltage potential V for an electrical flow through a homogeneous isotropic conducting

medium (*see* MAXWELL'S EQUATIONS) is likewise governed by an equation of the form of Eq (1) either the velocity potential or stream function or both can be related to the voltage through constants m and n called the scale factor

$$\text{Analogy A} \quad \varphi = mV$$
$$\text{Analogy B} \quad \psi = nV \qquad (3)$$

By use of the previous relations the fluid velocity components are proportional to the respective voltage gradients so that $u = m(\partial V/\partial x) = n(\partial V/\partial y)$ Every ideal fluid flow problem therefore has an electrical counterpart for this reason the electrolytic tank is also known as an electric tank analogy or potential flow analyzer Its solution includes construction of a scaled electrical flow model its installation in an electrical tank with proper simulation of the physical boundary conditions both on the model contour and on the field boundaries measurement of the electrical variables as required and finally the translation of the measurements by numerical computation into meaningful fluid flow terms

Apparatus For a typical two dimensional tank experience shows that slate is an excellent material for fabrication A reasonable size for a general purpose plane tank is 80 in long by 60 in wide by 5 in high Leveling screws must be provided to permit leveling of the tank bottom Cylindrical models 2–3 in in height of the desired cross section (such as an airfoil section) are sealed to the bottom

The tank is filled flush to the upper model surface with a weak electrolyte, ordinary tap water is often satisfactory as long as it does not react with the electrodes The electrodes are made of ¼ in brass plate and are clamped to opposite sides of the tank to establish a voltage field in problems involving a uniform free stream flow as illustrated

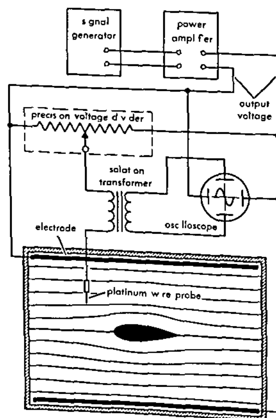
The electrical circuit includes a 1000-cps signal generator 20-watt power amplifier and oscilloscope used as a null indicator Special care must be used in the circuit to eliminate parasitical capacitance which could dephase the signal and obscure the true null

When it is desired to trace the potential field or voltage distribution along a given line the probe is supported by a carriage which can traverse the tank Attached to the carriage is a probe-follower In this manner any desired line can be followed and recorded relative displacements are measurable to ±0 01 in

Models Selection of the model material depends on the surface boundary condition and on the choice of analogy In the simplest case the model surface is a streamline Translated into electrical terms this means that in Analogy A the normal voltage gradient is zero and hence the model is a dielectric The surface in Analogy B must be a potential surface which is achieved by making it 'infinitely' conducting Plastics such as Lucite or Plexiglas make excellent nonconducting model plastic impregnated wood has also been used Conducting models can be machined from solid brass or copper stock or an overlay of brass sheet or sprayed silver base paint on a nonconducting form can be used

There are other boundary conditions on flows which require specialized techniques The Kutta Joukowski condition for an airfoil requires a flow with circulation such that a streamline leaves the trailing edge smoothly In Analogy A this is accomplished with a cut in the field with a constant voltage jump across the cut In Analogy B a supplementary current adjusted to give the desired streamline shape at the trailing edge is supplied to the model Techniques have also been devised to realize such diverse boundary conditions as finite sources or sinks boundary layer suction free streamlines and those of small perturbation theory for thin airfoils

Because a model must be constructed for each setup the analogy is not usually suited to general studies Its main advantage is that it can handle complex flow geometries which would be intractable to analysis and it is thus a useful engineering tool Problems successfully handled include lifting surface analysis inlet studies wind tunnel wall effects cascade flow wing body interference and nonstationary airfoil theory More sophisticated analogical applications include compressible flow



Setup for tracing streamlines of nonlifting airfoil Analogy B

conformal transformations, the hodograph and axisymmetric flows [W D BR]

Bibliography: L. Malavard, Electrolytic plotting tank *High Speed Aerodynamics and Jet Propulsion* vol 9, 1954, L. Malavard, *The Use of Rheoelectrical Analogies in Aerodynamics*, NATO Agardograph 18 1957.

Electromagnet

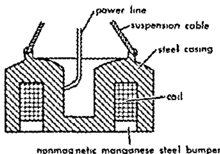
A soft iron core that is magnetized by passing a current through a coil of wire wound on the core Electromagnets are used to lift heavy masses of magnetic material and for attracting movable magnetic parts of electric devices, such as solenoids, relays and clutches

The difference between cores of an electromagnet and a permanent magnet is in the retentivity of the material used Permanent magnets initially magnetized by placing them in a coil through which current is passed are made of retentive (magnetically "hard") materials which maintain the magnetic properties for a long period of time Electromagnets are meant to be devices in which the magnetism in the cores can be turned on or off Therefore, the core material is nonretentive (magnetically "soft") material which maintains the magnetic properties only while current flows in the coil All magnetic materials have some retentivity called residual magnetism, the difference is one of degree See MAGNETIZATION

A magnet, when brought near other susceptible material induces magnetic poles in the susceptible material and so attracts it A force will be developed in the susceptible material that will tend to move it in a direction to minimize the reluctance of the flux path of the magnet The reluctance force may be expressed quantitatively in terms of the rate of change of reluctance with respect to distance See MAGNETISM

In an engineering sense the word electromagnet does not refer to the electromagnetic forces incidentally set up in all devices in which an electric current exists but only to those devices in which the current is primarily designed to produce this force as for instance in solenoids relay coils electromagnetic brakes and clutches and in tractive and lifting or holding magnets and magnetic chucks

Electromagnets may be divided into two classes traction magnets, in which the pull is to be exerted over a distance and work is done by reducing the air gap, and lifting or holding magnets in which the material is initially placed in contact with the magnet For examples of the first type see BRAKE, CLUTCH, RELAY, SOLENOID (ELECTRICAL) Examples of the latter type are magnetic chucks and circular lifting magnets The illustration shows a cross section view of a typical circular lifting magnet The outer rim makes up one pole and the inner area is the opposing pole Manganese steel used as a protective cover plate for the coil is nonmagnetic and does not provide a low reluctance shunt path for the flux.



Cross section of circular lifting electromagnet

The mechanical force between two parallel surfaces is given by Maxwell s equation

$$F = B^2A \quad 72\,13 \times 10^6 \quad \text{(lb)}$$

where B is the flux density (lines/in^2) and A is the cross sectional area (square inches) through which the flux passes When two poles are active the force produced by each is calculated to find the total force An interesting result of this relation is that the force is not simply the result of the total flux (BA) but also of the flux density Thus if the same flux can be forced through one half the cross sectional area the net pull will be doubled In practice it is difficult if not impossible to calculate the actual lifting capacity of the magnet using Maxwell's equation since the capacity varies with the shape and kind of material lifted how it is stacked, and other factors Therefore lifting magnets are usually rated on their all day average lifting capacity

Since currents are large (10–20 amp) and the circuit is highly inductive, control of a lifting magnet is a problem If the line switch were opened a destructive arc would result Therefore the controller employed with a lifting magnet usually does the following things automatically (1) reduces magnet current after initial high value to reduce heating of the magnet (2) introduces a shunt discharge resistor across the magnet before allowing the line to be opened when the operator turns the magnet off and (3) causes a reduced current of reverse polarity to flow in the magnet coil for a short time after the operator turns the switch off Thus the residual magnetism is cancelled and scraps and small chunks that might have continued clinging to the magnet will be released [J E G]

Bibliography A E Knowlton (ed), *Standard Handbook for Electrical Engineers*, 9th ed, 1957; E Molloy M G Say, and R C Walker (eds), *"Electrical Engineer" Reference Book*, 3d ed, 1948, H Pender and W A Del Mar, *Electrical Engineers' Handbook*, 4th ed, 1949

Electromagnetic field

A changing magnetic field always produces an electric field, and conversely, a changing electric field always produces a magnetic field This interaction of electric and magnetic forces gives rise to a re

gion in space known as an electromagnetic field The conditions in an electromagnetic field are expressed mathematically by the famous Maxwell equations *See* MAXWELL'S EQUATIONS, *see also* ELECTRIC FIELD, ELECTROMAGNETIC RADIATION, ELECTROMAGNETIC WAVE, MAGNETIC FIELD [K W P]

Electromagnetic propulsion

Motive power for flight vehicles produced by high speed discharge of a plasma fluid Together with electrostatic (ion) propulsion electromagnetic propulsion collectively designates several mechanisms capable of attaining specific impulses which exceed by one to two orders of magnitude those of thermal propulsion devices (Table 1) Much greater energy can be transferred to a body of matter by electrical means than by heating For a discussion of ionic propulsion, *see* ION PROPULSION The discharged plasma is electrically neutral (see MAGNETOGAS DYNAMICS, MAGNETOHYDRODYNAMICS) Discharge density is, therefore, not limited by electrostatic forces (space charge) present in an ion beam Electromagnetic propulsion devices offer promise of much higher thrust per unit discharge area and of operation inside the atmosphere as well as in space

Electromagnetic propulsion is adaptable to a wide range of specific impulses The optimization of a propulsion system within the framework of an over all vehicle system and a space mission calls for a compromise between the acceleration and the payload capability which best serves the particular mission purpose

The accelerator provides direction and speed to the plasma flow and thus represents the thrust producing mechanism Electromagnetic propulsion devices can be divided, on the basis of accelerating mechanism, into steady flow systems electropulsed systems, and magnetopulsed systems A survey of the presently recognized drives is presented in Table 2 Each drive is designated according to its main distinguishing characteristic for purposes of brevity No generally accepted nomenclature is available Many devices are closely related to each other

Partial plasma engine. The partial plasma engine applies intense arc heating and *partial* ionization of a fraction of the total discharge fluid (Fig 1) The current flowing through the plasma between the electrodes induces a *magnetic field*

Table 1 Comparison of specific impulses*

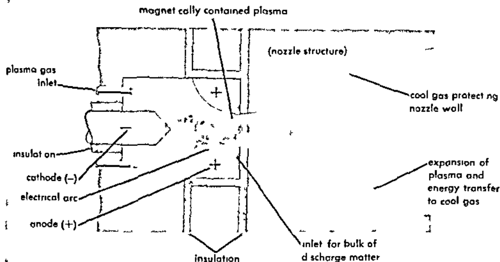
Drive	Specific impulse
Chemical	≤450
Solar heated	600–700
Nuclear heated	700–1 200
Electromagnetic	
Steady flow systems	1 400–5 000
Electropulsed systems	10 000–20 000
Magnetopulsed systems	≤10 000
Electrostatic	3 000–20 000
Fusion	100 000 (potentially)

* Specific impulse is the thrust force obtained per unit weight of fluid discharged per second (in consistent units) or briefly exhaust velocity divided by $g = 981$ cm/sec^2 = 32 2 ft/sec^2

Table 2 Survey of electromagnetic propulsion systems*

System	Designation	Accelerator	Propellant (working fluid)	Plasma formation	Potential specific impulse seconds (estimated)	Problems and remarks
Steady flow	Partial plasma engine	Magnetic compression and thermal expansion	Gas	Electric arc heating	1 400–2 000	Electrode erosion heat loss from neutral gas
Steady flow	Arc MHD engine	Thermal expansion augmented by emf	Gas	Electric arc heating	2 000–5 000	As above plus plasma stability unless operating at very low pressure potentially attractive but not yet sufficiently explored
Steady flow	RF MHD engine	As arc MHD engine	Gas	Radio-frequency heating	2 000–5 000	No electrode erosion possibly limited by heating efficiency potentially attractive but not yet sufficiently explored
Steady flow	Arc repulsor	Probably high intensity electrostatic repulsion	Electrode material	Electric arc heating	15 000	Insufficiently explored probably very low thrust per unit area
Electropulsed	Spark accelerator	High intensity emf	Gas metal	Spark discharge	10 000–20 000	Limitations imposed by condenser discharge
Electropulsed	MHD shock tube	Current loop expansion high intensity emf	Gas	Two-dimensional spark discharge	10 000–20 000	Like spark accelerator but higher gas densities appear possible attractive potential but not yet sufficiently explored
Electropulsed	MHD rotor accelerator shock tube	Current loop expansion in *strong* axial magnetic field	Gas	Two-dimensional spark discharge	10 000–20 000	As MHD shock tube
Magnetopulsed	Traveling wave accelerator	Traveling magnetic field	Gas	Arbitrary	≤10 000	Possible limitations due to charge separation tendencies 10 000 sec specific impulse demonstrated so far insufficiently explored

* RF = radio frequency MHD = magnetohydrodynamic emf = electromotive force



Fg 1 Partial plasma engine

which provides magnetic containment of the plasma As the plasma leaves the arc region it expands rapidly in the direction of the decreasing magnetic field and decreasing pressure transmitting its excess energy to the rest of the gas The process is followed by thermal expansion

High gas density is desired to establish high thrust density This requirement intensifies prob lems of heat transfer to the nozzle wall A high de gree of dissociation (high temperature and low pressure) reduces the gas molecular weight and yields higher specific impulse The plasma expan sion ratio can be increased with the aid of an ori ented external magnetic field

Arc MHD engine. In the arc magnetohydrody namic (MHD) engine over all gas flow is trans formed into a weakly ionized plasma (Fig 2) Thermodynamic expansion of the neutral gas is augmented by plasma acceleration through the electromagnetic force (emf) in crossed electric and magnetic fields To achieve plasma stability the electron cyclotron frequency (the product of electron charge e times magnetic flux density B divided by the product of 2π times electron mass m_e or $eB/2\pi m_e$) should be much larger than the particle collision frequency Thus either low fluid density or strong magnetic fields are required Steady fields of 10^4–10^5 gauss desirable in a gas pressure region of 10^{-3}–10^{-4} atm are required over extended periods of time

Arc heating methods suffer from limitations im posed by electrode erosion Erosion is avoided by radio frequency heating of the plasma However the efficiency of heating the gas tends to decrease as the plasma conductivity increases

Spark accelerator Electropulsed methods are based on the use of spark discharges through which intense electric and magnetic fields are established for periods of microseconds up to a few milliseconds A high voltage storage capacitor is discharged through a spark gap between two leads (rail electrodes) The intense current in the arc plasma is at right angle to the magnetic field which surrounds the current flow in the leads to the spark gap (Fig 3) The resulting emf drives the plasma along the leads and away from the spark gap

MHD shock tube In the MHD [illegible]

the current loop through the plasma expands with it for the duration of the discharge The high speed plasma expands into cooler gas in which a strong shock wave is formed causing intense heating and even ionization of the gas

The rotor accelerator is a modification in which a strong external axial magnetic field is superim posed by outside coils Interaction of this field and the radial current produces a tangential force accelerating the plasma between cylindrical elec trodes to high rotational velocity Upon expansion through the exhaust nozzle the rotational velocity is converted into axial motion The MHD rotor accelerator system offers longer stay time of the gas in the thrust chamber thus improving equilib rium conditions in the flow and permitting a more gradual energy flux into the working fluid The

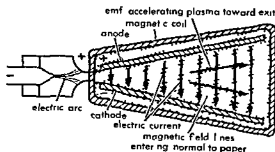


Fig 2 Arc MHD engine

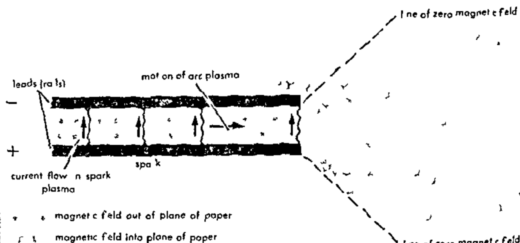


Fig 3 Spark accelerator

principal problem with all electropulsed methods is the limitation in present capacitor technology

The magnetopulsed system operates on the magnetic mirror effect A series of discrete coils is arranged along the plasma duct (each coil representing a converging diverging field pattern) in such a manner that the distance between coils increases in downstream direction The coils are sequentially energized by means of a multiphase radio frequency current The field appears to travel downstream at increasing velocity (traveling magnetic wave) Because of its diamagnetic characteristics the plasma in front of the magnetic wave is compressed and pushed forward at increasing speed Because of the mass difference between positive and negative charges this method may be limited by charge separation tendencies The resulting electrostatic field would produce an emf which is directed differently from the traveling wave force

Propellant fluids used in experimental plasma engines include water air argon helium hydrogen deuterium and metallic vapor Good potential working fluids are lithium some metal hydrides

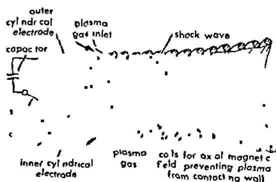


Fig 4 MHD shock tube

methane ammonia and hydrogen In principle plasma engines have the widest choice of propellants (at least from among the nonoxidizing fluids) of all propulsion systems This may eventually enable plasma driven space ships to draw their propellant from many celestial bodies rather than from Earth alone [K.A.E.]

Bibliography G Cann A Ducati and V Blackman Experimental studies on the thrust from a continuous plasma jet *Proc USAF OSR Rocketdyne Advanced Propulsion Symposium* 1957 K A Ehricke Comparison of propulsion systems *Proc USAF OSR Rocketdyne Advanced Propulsion Systems Symposium* 1957, J H Irving and E K Blum Comparative performance of ballistic and low thrust vehicles for flight to Mars *Second Annual USAF OSR Astronautics Symposium* 1958 W E Moeckel Propulsion methods in astronautics *First International Congress of the Aeronautical Sciences* 1958 R J Rosa Application of Magnetohydrodynamics to Propulsion *Proc USAF OSR Rocketdyne Advanced Propulsion Symposium* 1957

Electromagnetic pumps

Pumps that operate on the principle that a force is exerted upon a conductor (the fluid) carrying current in a magnetic field The high electrical conductivity of liquid metals (used as heat transfer media in some nuclear reactors) makes it possible to pump them by electromagnetic means For use in nuclear reactors where a minimal amount of maintenance is desirable electromagnetic pumps are often preferable to conventional mechanical pumps because they have no moving parts bearings or seals Various methods are employed to cause current to flow in the liquid metal

Direct current conduction pumps These pumps are a direct application of the right hand rule

which states that a current passing at right angles to a magnetic field will produce a force at right angles to both Pump performance depends upon the magnitude of the current magnetic field intensity and the geometry of the pump duct In its simplest form a pump of this type consists of a rectangular tube with electrodes attached to the short sides of the rectangular section and with the long axis of the section placed between the poles of a magnet Thus current flowing through the fluid along the long axis is cut by the magnetic field and produces a longitudinal thrust on the fluid in the tube Corrections must be made for the magnetic field produced by the flow of current through the duct walls and provision must be made to minimize end losses (flow of current through the fluid but outside the magnetic field) The disadvantage of this type of pump is the very high current (thousands of amperes) at low voltage (1-2 volts) required

Alternating current induction pumps Large currents can be developed in the liquid metal by electromagnetic induction An ac induction pump consists of a duct in the form of a flattened tube extending between two core sections containing a three phase ac winding The winding is similar to that of an induction motor stator except that the field structure is flat and a sliding rather than a *rotating* magnetic field is produced This pump employs conventional power supplies (60-cycle ac) but the field winding must be cooled to protect the electrical insulation

Other types of electromagnetic pumps have been developed and employed for laboratory use However the two types described here have received the most attention and their development has been carried to the most advanced levels including large commercial size units *See* PUMP REACTOR NUCLEAR [L J K]

Bibliography H Etherington (ed) *Nuclear Engineering Handbook,* 1958

Electromagnetic radiation

Energy transmitted through space or through a material medium in the form of electromagnetic waves The term can also refer to the emission and propagation of such energy Whenever an electric charge oscillates or is accelerated a disturbance characterized by the existence of electric and magnetic fields propagates outwards from it This disturbance is called an electromagnetic wave The frequency range of such waves is tremendous as is shown by the electromagnetic spectrum in the accompanying table The sources given are typical but not mutually exclusive as is shown by the fact that the atomic interstellar hydrogen radiation whose wavelength is 0 210614 m falls in the radar region The other monochromatic radiation listed is that from positron electron annihilation whose wavelength is $2\ 42626 \times 10^{-12}$ m

Detection of radiation In theory any electromagnetic radiation can be detected by its heating effect This method has actually been used over the range from x rays to radio Ionization effects measured by cloud chambers photographic emulsions ionization chambers and Geiger counters have been used in the γ and x ray regions Direct photography can be used from the γ ray to the infrared region Fluorescence is effective in the x ray and ultraviolet ranges Bolometers thermocouples and other heat measuring devices are used chiefly in the infrared and microwave regions Crystal detectors vacuum tubes and transistors cover the microwave and radio frequency ranges

Free space waves A charge in simple harmonic (linear sinusoidal) motion in a vacuum generates a simple wave which becomes spherical at distances from the source much larger than the amplitude of

Electromagnetic spectrum

Frequency cps	Wave length m	Nomenclature	Typical source
10^{23}	3×10^{-15}	Cosmic photons	Astronomical
10^{22}	3×10^{-14}	γ rays	Radioactive nuclei
10^{21}	3×10^{-13}	γ rays x rays	
10^{20}	3×10^{-12}	x rays	Atomic inner shell
		Positron electron annihilation	
10^{19}	3×10^{-11}	Soft x rays	Electron impact on a solid
10^{18}	3×10^{-10}	Ultraviolet x rays	Atoms in sparks
10^{17}	3×10^{-9}	Ultraviolet	Atoms in sparks and arcs
10^{16}	3×10^{-8}	Ultraviolet	Atoms in sparks and arcs
10^{15}	3×10^{-7}	Visible spectrum	Atoms hot bodies molecules
10^{14}	3×10^{-6}	Infrared	Hot bodies molecules
10^{13}	3×10^{-5}	Infrared	Hot bodies molecules
10^{12}	3×10^{-4}	Far infrared	Hot bodies molecules
10^{11}	3×10^{-3}	Microwaves	Electronic devices
10^{10}	3×10^{-2}	Microwaves radar	Electronic devices
10^{9}	3×10^{-1}	Radar	Electronic devices
		Interstellar hydrogen	
10^{8}	3	Television FM radio	Electronic devices
10^{7}	30	Short wave radio	Electronic devices
10^{6}	300	AM radio	Electronic devices
10^{5}	3000	Long wave radio	Electronic devices
10^{4}	3×10^{4}	Induction heating	Electronic devices
10^{3}	3×10^{5}		Electronic devices
100	3×10^{6}	Power	Rotating machinery
10	3×10^{7}	Power	Rotating machinery
1	3×10^{8}		Commutated direct current
0	Infinity	Direct current	Batteries

the motion and so great that many oscillations have occurred before the disturbance arrives The wave is plane when the dimensions of the area observed are very small compared with the radius of spherical curvature In this case the choice of the rectangular coordinates x and z as the directions of the oscillation and of the observation or field point respectively permits the electric intensity **E** and the magnetic flux density **B** to be written

$$E_x = \iota B_y - E_0 \cos [\omega(t - v^{-1}z)] \qquad (1)$$

The field amplitude E_0 is constant over the specified area and not dependent on z if the z range is small compared with the source distance as in stellar radiation The angular frequency of the source is ω radians per second which is the frequency ν in cycles per second multiplied by 2π The velocity of the wave is v the direction of propagation z and the time t The wavelength λ is $2\pi v/\omega$ If t is in seconds and z in meters then v is in meters per second and λ in meters It is found that in a lossless isotropic homogeneous medium

$$v = (\mu\epsilon)^{-1/2} \qquad (2)$$

where μ is the permeability and ϵ the capacitivity or dielectric constant This wave is transverse because **E** and **B** are normal to z It is plane polarized because E_x and B_y are parallel to fixed axes The plane of polarization is taken as that defined by the electric vector and the direction of propagation

Plane waves An electromagnetic disturbance is a plane wave when the instantaneous values of any field element such as **E** and **B** are constant in phase over any plane parallel to a fixed plane These planes are called wavefronts In empty unbounded space **E** and **B** lie in the wavefront normal to each other if the wave is unpolarized their direction fluctuates in this plane in random fashion If the plane waves are bounded as on transmission lines and in wave guides the amplitudes may vary over the wavefront and in the case of wave guides and crystals some of the elements will not in general lie in the wavefront The equation for an undamped plane wave whose front is normal to z is

$$F = \Phi_1(x\,y) f_1(z - \iota t) + \Phi_2(x\,y) f_2(z + \iota t) \quad (3)$$

where F is one of the field elements such as **E** or **B** Note that if an observer sees a certain value of $\Phi_1(x\,y)$ at z and then jumps instantaneously in the z direction to a point $z + \Delta z$ he will after waiting a time $\Delta z/v$ see the same value $\Phi_1(x\,y)$ because

$$f(z - \iota t) = f[z + \Delta z - v(t + \Delta z/v)]$$

Thus the first term represents a wave moving in the z direction with a velocity ι The second term represents a wave in the negative z direction The form of $\Phi_1(x\,y)$ and $\Phi_2(x\,y)$ depends on the boundary conditions *See* WAVE EQUATION

Spherical waves A wave is spherical when the instantaneous value of any field element such as **E** or **B** is constant in phase over a sphere The radiation from any source of finite dimensions becomes spherical at great distances in an unbounded isotropic, homogeneous medium The equation for an undamped spherical wave is

$$F = r^{-1}\Phi_1(\theta,\varphi) f(r - \iota t) + r^{-1}\Phi_2(\theta\,\varphi) f(r + \iota t) \quad (4)$$

The first term represents a diverging and the second a converging wave Again the form of $\Phi_1(\theta\,\varphi)$ and $\Phi_2(\theta\,\varphi)$ depends on the nature of the source and other boundary conditions

Damped waves If there are energy losses which are proportional to the square of the amplitude as in the case of a medium of conductivity γ which obeys Ohm's law then the wave is exponentially damped and Eq (1) becomes

$$E_x = E_0 e^{-\alpha z} \cos (\omega t - \beta z) \qquad (5)$$

The symbol α is called the attenuation constant and β the wave number or phase constant which equals ω/v' where v' is the damped wave velocity The electric wave amplitude at the origin has been taken as E_0 The ratio of E_0 to B_0 as well as that of α to β depends on the permeability μ the capacitivity ϵ and the conductivity γ of the medium In terms of the phasor $\check{E}_x$, Eq (5) may be written as the real part of

$$E_x = \check{E}_x e^{j\omega t} = E_0 e^{-(\alpha + j\beta)z} e^{j\omega t} \qquad (6)$$

This is exactly the form for the current on a transmission line (Phasors are complex numbers of form such that when multiplied by $e^{j\omega t}$, the real part of the product gives the amplitude phase, and time dependence)

Wave impedance Those trained in transmission line theory find it useful to apply the same techniques to wave theory Consider an isolated tubular section of the wave in Eq (1) bounded by $x = 0$ $x = 1$ and $y = 0$ $y = 1$ as a transmission line The potential across the line between $x = 0$ and $x = 1$ is **E** The line integral of **B** around the $x = 0$ boundary from $y = 0$ to $y = 1$ is μI by Ampere's law and equals **B** because **B** is zero on the negative side Thus the impedance of the line is making use of Eqs (1) and (2),

$$\check{Z}_k = \frac{V}{I} = \frac{\mu E}{B} = \frac{E}{H} = \left(\frac{\mu}{\epsilon}\right)^{1/2} = \eta \qquad (7)$$

This depends only on the properties of the medium and is known as the wave impedance In transmission line theory, the ratio μ/ϵ would be replaced by the ratio of the series impedance $\check{Z}_L = j\omega L$ to the shunt admittance $\check{Y} = j\omega C$ where L is the inductance per unit length and C the capacitance per unit length across the line If there is a resistance R per unit length across the line then $1/R$ must be added to $\check{Y}$ This resistance is $1/\gamma$ for the tubular section Thus for a conducting medium Eq (7) becomes

$$\check{Z}_k = \left(\frac{j\omega\mu}{\gamma + j\omega\epsilon}\right)^{1/2} = \frac{j\omega\mu}{\alpha + j\beta} \qquad (8)$$

Electric field lines generated by Hertzian oscillator shown at eighth-period intervals (a) $t = 0$ (b) $t = T/8$ (c) $t = T/4$ (d) $t = 3T/8$

The last term is a common transmission line form The reflection and refraction of plane waves at plane boundaries separating different media may be calculated by transmission line formulas with the aid of Eqs (7) and (8) *See* REFLECTION AND . . .

n
tl
t
F

shaped conductor in which the electrons oscillate from one end to the other, leaving the opposite end periodically positive An electric dipole of moment M is defined as the product qa when two large equal and opposite charges, $+q$ and $-q$, are placed a small distance a apart A dipole is oscillating when M is periodic in time and is the simplest source of spherical waves Much can be learned by a study of H Hertz's picture of the outward moving electric field lines at successive time intervals of one-eighth period in a plane which passes through the Hertzian oscillator axis shown in the figure The most striking feature of the pictures is that after breaking loose from the dipole all electric field lines are closed, which means that the divergence of E is zero This is true of all unbounded waves It is also noteworthy that the waves become truly spherical with a fixed wavelength λ only in a direction perpendicular to the dipole and at a distance which greatly exceeds the dipole dimensions This distance is beyond the edges of the picture Lengths $\lambda/4$ $\lambda/2$ and $3\lambda/4$ are marked off on the axis for comparison The magnetic field lines are circles coaxial with the oscillator, so they intersect the plane of the diagram normally They are most dense where the electric lines are closely spaced The radiant energy emitted by atoms and molecules is essentially radiation of the dipole type *See* ABSORPTION (ELECTROMAGNETIC RADIATION), ANTENNA (AERIAL), DIFFRACTION, GAMMA RAYS, HEAT RADIATION, INFRARED RADIATION, INTERFERENCE OF WAVES, LIGHT, MAXWELL'S EQUATIONS, MICROWAVE, POLARIZATION OF WAVES, RADIATION, RADIO WAVE PROPAGATION, REFLECTION (ELECTROMAGNETIC RADIATION), REFRACTION OF WAVES,

SCATTERING (ELECTROMAGNETIC RADIATION) TRANSMISSION LINES TRANSMISSION THEORY AND METHODS ULTRAVIOLET RADIATION WAVE GUIDE WAVE MOTION X RAY(S) PHYSICAL NATURE OF [W P SM]

Bibliography M Born and E Wolf *Principles of Optics* 1959 A H Compton and S K Allison *X Rays in Theory and Experiment* 1935 G P Harnwell *Principles of Electricity and Electromagnetism* 2d ed 1949 F A Jenkins and H E White *Fundamentals of Optics* 3d ed 1957 L Page and N I Adams *Principles of Electricity* 3d ed 1958 S Ramo and J R Whinnery *Fields and Waves in Modern Radio* 1953 S A Schelkunoff *Electromagnetic Waves* 1943

Electromagnetic wave

A disturbance produced by the acceleration or oscillation of an electric charge which has the characteristic time and spatial relations associated with progressive wave motion A system of electric and magnetic fields moves outward from a region where electric charges are accelerated such as an oscillating circuit or the target of an x ray tube The wide wavelength range over which such waves are observed is shown by the electromagnetic spectrum (*see* ELECTROMAGNETIC RADIATION) The term electric wave or Hertzian wave is often applied to electromagnetic waves in the radar and radio range Electromagnetic waves may be confined in tubes such as wave guides or guided by transmission lines They were predicted by J C Maxwell in 1864 and verified experimentally by H Hertz in 1884 [W R SM]

Electromagnetism

The branch of science dealing with the observations and laws relating electricity to magnetism Electromagnetism is based upon the fundamental observations that a moving electric charge produces a magnetic field and that a charge moving in a magnetic field will experience a force

The magnetic field produced by a current is related to the current the shape of the conductor and the magnetic properties of the medium around it by Ampere s law *See* AMPERE S LAW

The magnetic field at any point is described in terms of the force that it exerts upon a moving charge at that point The electrical and magnetic units are defined in terms of the ampere which in turn is defined from the force of one current upon another See AMPERE

The association of electricity and magnetism is also shown by electromagnetic induction in which a changing magnetic field sets up an electric field within a conductor and causes the charges to move in the conductor *See* INDUCTION ELECTROMAGNETIC *see also* EDDY CURRENT ELECTRICITY ELECTROMAGNET FARADAY S LAW OF INDUCTION HALL EFFECT INDUCTANCE INDUCTION MAGNETIC LENZ S LAW MAGNETIC CIRCUITS MAGNETIC FIELD MAGNETIC FLUX MAGNETISM MAGNETOMOTIVE FORCE, MAXWELL'S EQUATIONS RELUCTANCE [K V M]

Electrometallurgy

That portion of process metallurgy in which an electric current is utilized to bring about a purification of the metal (electrorefining) or to reduce a metallic compound to the metal (electrowinning)

Electrorefining This is a purification process in which the impure metal is made the anode (positive electrode) in a solution of a salt of the metal being refined The pure metal deposits at the cathode during electrolysis

Electrorefining has proved to be the most economical method for securing the high purity required for many commercial nonferrous metal This process was applied to copper in the latter part of the nineteenth century or as soon as the electric generator was invented Electrorefined copper is an important material in the electrical industry since minor quantities of some impurities will lower the electrical conductivity of copper very markedly for example 0 01% arsenic in copper lowers the conductivity by 3% Further silver and gold are common constituents of copper ore and follow along with copper through all the pyrometallurgical steps The removal of the silver and the gold from the copper is carried out in the electrolytic refining step Thus the refining process ensures that the metal will meet the specification of the purchaser and permits the recovery of precious metal impurities

Impure metal slabs are cast in varying sizes and shapes depending upon the particular metal being refined to permit vertical hanging in the refining cell In the case of copper the anode is approximately 36 in by 36 in by 1½ in cast with lugs to support the anode from the walls of the cell or from the bus bars These anodes weigh about 650–700 lb each The cathodes are made of pure copper sheet usually deposited on a smooth starting blank stripped from the blank and then suspended from a copper bar The electrode spacing from anode center to anode center is approximately 4 in A current of about 15–25 amp/ft of electrode surface is passed This will consume about 20–30 lb of anode per day depending on current density current efficiency and the amount of slime formed The solution employed contains a salt of the metal being refined In the case of copper a solution containing about 45 g/liter of copper as copper sulfate with approximately 200 g/liter of sulfuric acid is used A temperature of 55–60°C is maintained to lower the resistance and the electrolyte is circulated through the cell The size of the cell is again dependent upon the metal being refined In the case of copper the tanks are 12 14 ft long about 3½ ft wide and 3¾ ft deep This allows for approximately 30–35 anodes and one more cathode than anode in each cell In those cases where a sulfuric acid electrolyte is used lead lined wooden or concrete tanks are employed All pipes and pump



Lifting load of electrorefined copper from refining cell (Anaconda Copper)

are lead or lead lined Where the electrolyte is a fluosilicate as in the case of lead refining hard rubber pipes and pumps and asphaltum lined concrete tanks are used The electrical connections are made such that the electrodes in each cell are in parallel with each other and cells are placed in series

In the refining process purification is accomplished at both the anode and the cathode Metals such as copper silver gold and lead exhibit very little irreversibility and therefore the electrical potential at which they will dissolve is fairly close to the reversible potential which is given by the equation

$$E = E^0 - \frac{RT}{nF} \ln a_m$$

where E is the equilibrium or reversible potential of the electrode E^0 is the standard electrode potential for the metal R is the gas constant F is the value of the faraday T is the absolute temperature n is the valence charge during electrolysis and a_m is the activity (effective concentration) of the metal ion in solution (*see* ELECTRODE POTENTIAL) There is some polarization of the electrodes during electrolysis which increases as the current density is increased so that the anode potential must be slightly more anodic than the equilibrium potential in order for the reaction to proceed at an economical rate

The activity of the ion is in all cases considerably less than one When the above equation is applied in turn to all of the metals appearing in the anode as impurities it can be shown that some of the metals dissolve along with the metal being refined while others will not go into solution because the potential is not sufficiently anodic Again taking the case of copper such metals as iron nickel and lead which are in the anode copper will dissolve The lead does not stay in solution since the solubility product of lead sulfate will be exceeded and the lead will appear as lead sulfate in the anode slimes Such metals as gold and silver will not dissolve at the potential the anode assumes in copper refining but will remain as metals in the anode slimes By far the highest percentage of material in the anode slimes is copper in the form of particles that remain from the anodic solution of the bulk of the metal These are not dissolved possibly due to crystal orientation or to the proximity in the anode of more noble metals such as gold and silver These slimes are processed further for recovery of the various metals that they contain

Since the electrode processes being discussed are nearly reversible the cathode potential need be only slightly more cathodic than the equilibrium potential calculated by the above equation Under these conditions metals less noble than copper will not deposit whereas metals more noble than copper would deposit if they were in solution However as discussed under the anodic process the more noble metals do not dissolve Thus the nickel which did dissolve from the anode will not deposit at the cathode but the solution continues to build up in nickel concentration This necessitates periodic purification of the electrolyte to remove the nickel salts which continually build up

Nickel does not behave reversibly at an electrode and therefore difficulties are encountered in nickel refining since copper and nickel both dissolve at the anode potential required for the nickel purification process to take place at an economical rate For a discussion of these problems *see* NICKEL

Special agents are added to the electrolytes in most electrorefining processes to control the physical properties of the cathode deposits

Electrowinning This process is an electrochemical reduction of a metallic compound to the metal In this process the ore or a roasted concentrate is leached with an acid solution The solution is then circulated through a cell in which are suspended insoluble anodes and cathodes High purity metal can be obtained by this method provided that proper solution control is maintained In the case of zinc winning the electrolyte is a sulfuric acid zinc sulfate solution in a lead lined concrete tank The anodes are lead containing small amounts of silver usually have about 9 ft^2 of submerged surface and weigh about 100 lb each Aluminum sheet is usually the cathode material The accumulated zinc is stripped from the cathode sheet approximately once a day

A calcine which is soluble in sulfuric acid is produced by roasting an ore of the metal Impurities originally present on the ore that are soluble in sulfuric acid are removed in two steps The first involves increasing the pH (lowering the acidity) of the leach solution to the point where the ferric hydroxide precipitates This precipitate has a strong tendency to adsorb and coprecipitate many impurities The electrolyte is next treated with zinc dust and metals more noble than zinc are galvanically replaced that is precipitated from the solution It is imperative that all metals with low hy

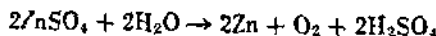
drogen overvoltage (*see* OVERVOLTAGE) be removed from the solution before the electrolyte is introduced into the cell Traces of germanium will result in practically zero cathode current efficiency Current efficiency is the ratio of the amount of electricity (coulombs) theoretically required to yield a given quantity of metal to the amount actually consumed Many other impurities act the same way but not all of them as markedly as germanium. Since the reduction potential of zinc is far more cathodic than that of hydrogen ion the cathodic deposition of zinc depends upon maintaining a high hydrogen overvoltage Metals of low hydrogen overvoltage remaining in the electrolyte will deposit if they are more noble than zinc forming centers for hydrogen evolution on the cathode and thereby reducing the current efficiency for zinc deposition The presence of these more noble metals on the cathode also promotes the dissolution of zinc already deposited by setting up a local galvanic cell the cathodic potential being insufficient to yield complete cathodic protection this again reduces the current efficiency Thus in practice great care is exercised to see that the solutions are purified extensively These difficulties are not nearly as serious in the case of electrowinning of more noble metals such as copper

Whereas in electrorefining it is normal to have cell voltages of the order of 0 2 volt in electrowinning cell voltages of the order of 2 3½ volts are common The reason for this large difference in electrical *power consumption* lies in the type of chemical changes produced The weight of metal *deposited as given* by Faraday s laws is a function of *the number* of ampere hours passed and the equivalent weight of the metal while the power *required* for the process is a function of these two factors and *the voltage* drop across the cell In the case of *electrolytic refining* very little chemical work is accomplished The metal dissolves from an impure anode *and deposits* on a pure cathode The free energy *change for* this process is negligible In the case of electrowinning *of zinc for* example the net chemical reaction taking place in the cell is

$$2ZnSO_4 + 2H_2O \rightarrow 2Zn + O_2 + 2H_2SO_4$$

The free energy change accompanying this reaction is considerable *and is related* to the reversible voltage by *the equation* $-\Delta G = nFE$ where ΔG is the Gibbs free energy *change* n is the valence change F is the Faraday constant and E is the voltage required to *drive the reaction*

Many metals less noble *than* zinc cannot be cathodically reduced from aqueous solutions In the case of these metals it is standard to



fused salt electrolysis With *the exception* of aluminum the majority of these metals *are prepared* by the electrolysis of their fused chlorides usually with the addition of other salts *to lower* the melting point of the electrolyte sufficiently that the vapor pressure of the metal being produced is low enough to make it possible to recover the product economically Many of these metals are of lower specific gravity than the electrolyte from which they are produced and therefore float to the surface Care must be exercised to protect them from coming in contact with the air or spontaneous combustion will result Further the cells must be designed so that the product of the anodic reaction does not come in contact with the molten metal When a fused chloride is being electrolyzed the anodic product is chlorine gas which is always recovered In the case of aluminum the usual elec

cally The anodes are carbon modern p... the Soderberg electrode in which the carbon electrode is formed in place from pitch and coke thus obviating the need for a separate electrode manufacturing plant The cells are constructed of steel lined with carbon brick The cell lining acts as cathode Again typical of electrowinning the cell voltages are *high—of the order* of 3-6 volts *See* [H B L.]
ELECTROCHEMISTRY, PYROMETALLURGY

Bibliography C. L. Mantell *Industrial Electrochemistry* 3d ed 1950

Electrometer

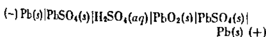
A highly sensitive instrument used to measure a voltage without drawing appreciable current by measuring the electrostatic force exerted between two bodies that are charged with the voltage In one form two suspended parallel strips of gold leaf spread out at an angle proportional to the voltage to which they are charged the amount of movement is measured with a microscope having a calibrated scale In a string electrometer the gold strips are replaced by lightly stretched metallized quartz fibers Electrometers involving mechanical movements are largely being replaced today by vacuum tube electrometers which are essentially voltage measuring amplifiers having such a high input resistance (usually above 10 ⁰ ohms) that they draw practically no current *See* ELECTROSCOPE VOLTAGE MEASUREMENT VOLTMETER [J MB]

Electromotive force (cells)

When two dissimilar electrodes are connected through an external conducting circuit a difference in electric potential exists between them *Although* this difference in potential is sometimes *called the* potential difference of the electrode *couple* it is customary to say that the galvanic cell *composed* of the two dissimilar electrodes exhibits an electromotive (driving) force This electromotive force (emf) is the resultant of the relative potentials forces of the two dissimilar electrodes at which electrochemical reactions occur during cell operation The two dissimilar electrodes need not be of unlike metals for example the metals may both

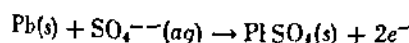
be copper but with the two coppers immersed in solutions of different concentration or composition Likewise both electrodes may be of the same gas but with the pressure of the gas different at the two electrodes See ELECTRODE POTENTIAL

Cell types Galvanic cells are of two general types reversible and irreversible If the chemical reactions at the electrodes can be exactly reversed by reversals in the direction of the current flow at the electrodes the cell is said to be reversible if the chemical reactions cannot be reversed or if entirely different reactions occur on current reversal the cell is then of the irreversible type An example of a reversible cell is

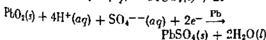


$$(-)\,Pb(s)|PbSO_4(s)|H_2SO_4(aq)|PbO_2(s)|PbSO_4(s)|Pb(s)\ (+)$$

where s = solid aq = aqueous solution and the vertical lines represent the interfaces between different pha es This cell is the familiar lead acid storage cell (or battery), widely used for a variety of purposes including starting ignition and lighting for automobiles When the cell is discharged that is when electric energy is being drawn from it the following reactions occur at the negative and positive electrodes respectively

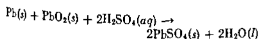


$$Pb(s) + SO_4^{--}(aq) \rightarrow PbSO_4(s) + 2e^-$$



$$PbO_2(s) + 4H^+(aq) + SO_4^{--}(aq) + 2e^- \xrightarrow{Pb} PbSO_4(s) + 2H_2O(l)$$

Here the symbol l = liquid The over all cell reaction is

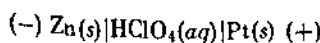


$$Pb(s) + PbO_2(s) + 2H_2SO_4(aq) \rightarrow 2PbSO_4(s) + 2H_2O(l)$$

whereby lead sulfate and water are formed in the cell reaction The lead in the reaction at the positive electrode merely serves as an electronic conductor

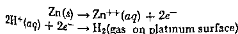
Now when the cell is charged the reverse of the above reactions occurs and lead sulfate is converted back to the initial state The cell is therefore called a reversible one

If however a cell is prepared by immersing zinc and platinum electrodes in perchloric acid an irreversible cell results This cell may be represented as



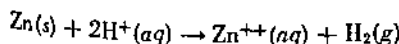
$$(-)\,Zn(s)|HClO_4(aq)|Pt(s)\ (+)$$

When the cell is discharged the following reactions occur at the negative and positive electrodes respectively



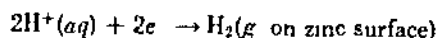
$$Zn(s) \rightarrow Zn^{++}(aq) + 2e^-$$
$$2H^+(aq) + 2e^- \rightarrow H_2(\text{gas on platinum surface})$$

The over all cell reaction is



$$Zn(s) + 2H^+(aq) \rightarrow Zn^{++}(aq) + H_2(g)$$

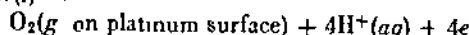
Here g = gas Now if the cell is charged that is subjected to an electrolyzing current instead of discharged the reactions are



$$2H^+(aq) + 2e \rightarrow H_2(g \text{ on zinc surface})$$

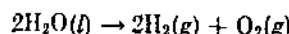


$$2H_2O(l) \rightarrow$$



$$O_2(g \text{ on platinum surface}) + 4H^+(aq) + 4e$$

at the negative and positive electrodes respectively The over all cell reaction is then



$$2H_2O(l) \rightarrow 2H_2(g) + O_2(g)$$

or the simple electrolysis of water Since the electrode reactions at each electrode for the charge differ from those obtained for the discharge the cell is of the irreversible type

In many cases the reversibility of a cell or of the electrodes comprising the cell can be determined only by means of precise electrical measurements Reversibility in the strict sense is determined by the response of a cell to very small (infinitesimal) discharging or charging current In practice measurements are made simultaneously of the current and the electromotive force of the cell for different values of the current The slope of the electromotive force of the cell versus the current is a certained for both the charging and the discharging currents For reversible cells the two slopes should be identical furthermore these slopes should be reproducible for repeated reversals in the direction of the flow of current through the cell Also for reversible cells the internal re

[illegible]

fected by the passage of very small currents through it in either direction The reversibility of a galvanic cell is primarily a function of the reversibility of the two electrodes comprising the cell and the same criteria may be used to establish the re

[illegible]

achieved by balancing the electromotive force of the cell against the electromotive force of another cell of known and steady value using a highly sensitive galvanometer When this state is achieved

[illegible]

lombs corresponding to the cell reaction it gives

[illegible]

valence change) involved in the cell reaction F is the faraday and ΔG is the change in (Gibbs) free energy associated with the cell reaction Furthermore if the variation of the electromotive force of the cell with temperature is measured the heat of reaction ΔH for the cell may be computed by the

well known Gibbs Helmholtz equation

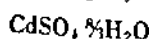
$$\Delta H = -nFE + nFT\frac{dE}{dT}$$

which may also be written as $\Delta H = \Delta G + T\Delta S$ where ΔS is the entropy change for the cell reaction since $nF(dE/dT) = \Delta S$

If a galvanic cell has a negative emf temperature coefficient the heat of the reaction exceeds the free energy change thus the available electrical work is less than that corresponding to the heat of the reaction for the cell the cell warms in operation and heat is lost to the surroundings Conversely if the emf temperature coefficient is positive the free energy exceeds the heat of reaction and the cell tends to cool when in operation this cooling is overcome if the cell is maintained under isothermal conditions by absorption of heat from the surroundings thus the total available electrical energy from such a cell is a resultant of the inherent changes in the heat content of the cell and the heat absorbed from the surroundings *See* FREE ENERGY

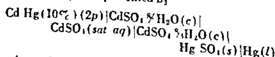
Electromotive force measurements As stated above in measuring the emf of a cell a reference cell of known emf must be available to effect a comparison This reference cell should have an emf that is known in terms of physical laws and units and not one chosen arbitrarily The emf of reference cells (and then through comparisons the emf of all cells) is known in terms of the cgs (centimeter gram second) system of electromagnetic units through Ohm s law $E = IR$ where E is emf in volts I is current in amperes and R is resistance in ohms In practice then the emf of a cell is equal to IR and it may be balanced against the IR drop across a resistor of known value through which a known current is flowing If the values of the resistor and the current are known in cgs electromagnetic units then the emf of a cell is given in like units *See* POTENTIOMETER (VOLTAGE MEASUREMENT)

Standard cells The standard cells used for this purpose are the cadmium amalgam standard cells of the saturated type proposed by Edward Weston in 1892 This type of cell is the most reversible galvanic cell known and retains a constant emf to within a few microvolts for many decades This cell consists of a cadmium amalgam anode (negative element) a mercury mercurous sulfate cathode (positive element) and a saturated solution of cadmium sulfate containing crystals of



$$CdSO_4 \cdot \tfrac{8}{3}H_2O$$

This cell may be represented by



$$Cd\,Hg(10\%)(2p)|CdSO_4 \cdot \tfrac{8}{3}H_2O(c)|\;CdSO_4(sat\ aq)|CdSO_4 \cdot \tfrac{8}{3}H_2O(c)|\;Hg_2SO_4(s)|Hg(l)$$

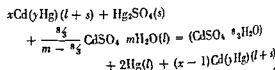
where $2p$ = two phase c = crystals $sat\ aq$ = saturated aqueous solution and the other symbols have the meaning given above This cell has an emf when freshly made of 1.018636 volts at 20°C and for precise measurements must be maintained at a constant temperature Its emf at other temperatures between 0 and 43.5°C may be calculated from the equation

$$E_t = 1.018636 - 0.0000406(t - 20) - 0.00000095(t - 20)^2 + 0.00000001(t - 20)^3$$

where t is in degrees centigrade Saturated standard cells should not be used above 43.5°C at this temperature the crystals of $CdSO_4 \cdot \frac{8}{3}H_2O$ are converted to $CdSO_4 \cdot H_2O$ Although standard cells prepared with the monohydrate are stable they can be used with confidence only at temperatures above 43.5°C, when such cells are cooled the monohydrate reverts to $CdSO_4 \cdot \frac{8}{3}H_2O$ but the rate of conversion is slow and the emf of such cells is erratic for indefinite periods

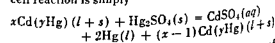
The cadmium standard cell is also made in the unsaturated type that is with no crystals of $CdSO_4 \cdot \frac{8}{3}H_2O$ and is the type widely used in recording instruments with potentiometers and in pH meters A solution of cadmium sulfate that is saturated at 4°C is used in its preparation the solution is then unsaturated at higher and normal temperatures It is made portable by placing cork or plastic septa over the positive and negative elements This cell has a very low emf temperature coefficient (0.000005 volt per °C), about one-tenth that of the saturated type On the average this cell decreases in emf at the rate of 70 microvolts per year and its ultimate life is about 10 years

For the saturated type of cell the over all reaction of the cell is



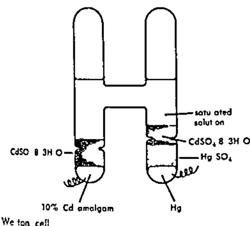
$$xCd(yHg)(l+s) + Hg_2SO_4(s) + \frac{8/3}{m - 8/3}CdSO_4 \cdot mH_2O(l) = (CdSO_4 \cdot \tfrac{8}{3}H_2O) + 2Hg(l) + (x-1)Cd(yHg)(l+s)$$

where x moles of Cd are associated with y moles of Hg in the amalgam and m is the number of moles of water associated with 1 mole of $CdSO_4$ in the saturated solution For the unsaturated cell the cell reaction is simply



$$xCd(yHg)(l+s) + Hg_2SO_4(s) = CdSO_4(aq) + 2Hg(l) + (x-1)Cd(yHg)(l+s)$$

A cross sectional sketch of the saturated type of cell is shown in the illustration The unsaturated type is made similarly except that no crystals are used and cork or plastic septa are placed above the amalgam and the mercurous sulfate paste The H form container is made of Kimball glass with platinum wires sealed in the bottom of each limb Mercury purified by several vacuum distillations is placed at the bottom of one limb and is used to prepare the amalgam used in the bottom of the other limb The cadmium is purified by sublimation A 10% (sometimes 12.5%) amalgam of cadmium is added while warm and in a single phase on cooling it becomes two phased with the solid phase being an isomorphous mixture of cadmium and mercury Mercury(I) sulfate prepared electrolyti



We ton cell

cally is then placed over the mercury and crystals of $CdSO_4$ ⅝H_2O are added to both limbs Then a saturated solution of cadmium sulfate is added to a le el slightly above the crossarm and the cell is sealed The unsaturated cell is usually mounted in a nontransparent case and the saturated cell is housed in oil baths or in thermoregulated air baths After proper aging these cells serve as reliable standards of emf with which the emf of all other cells are compared *See* BATTERY (ELECTRIC) CALOMEL ELECTRODE ELECTROCHEMISTRY GLASS ELECTRODE HYDROGEN ELECTRODE SALT BRIDGE [W J H]

Bibliography P Delahay *New Instrumental Methods in Electrochemistry* 1954

Electromotive force (emf)

The electromotive force $\mathcal{E}$ around a closed path in an electric field is the work per unit charge required to carry a small positive charge around the path It may also be defined as the line integral of the electric intensity around a closed path in the field The abbreviation emf is preferred to the full expression since emf also called electromotance is not really a force The term emf is applied to sources of electric energy such as batteries generators and inductors in which current is changing

The magnitude of the emf of a source is defined as the electrical energy converted inside the source to some other form of energy exclusive of electrical energy converted irreversibly into heat or the amount of some other form of energy converted in the source into electrical energy when a unit charge flows around the circuit containing the source In an electric circuit except for the case where an electric current is flowing through resistance and thus electrical energy is changed irreversibly into heat energy electrical energy is converted into another form of energy only when current flows against an emf On the other hand some other form of energy is converted into electrical energy only when current flows in the same sense as an emf

For a discussion of motional emf *see* INDUCTION ELECTROMAGNETIC For information on back or counter emf *see* DIRECT CURRENT MOTOR *See also* ELECTROMOTIVE FORCE (CELLS) MAGNETOMOTIVE FORCE [R P WI]

Bibliography F W Sears *Electricity and Magnetism* 1951

Electromyogram

A record of the electrical response of a muscle to stimulation Muscular contraction results from the rhythmical discharge of motor nerve cells each of which controls a group of 10–500 muscle fibers known as a motor unit The intensity of contraction reflects the number of units activated among the many hundreds comprising a muscle Each nerve impulse in a series sets up a brief cycle of contraction in the corresponding muscle fibers beginning with a depolarization and corresponding electrical change the muscle action potential The electrical change in muscle resulting from a single nerve impulse is a diphasic variation of 100–500 microvolts lasting 20–60 milliseconds and depending upon the size and grouping of the unit fibers A willed contraction is seen as a profusion of such potentials (interference pattern) from the unrelated rhythms of many units near the recording electrode Some conditions synchronize the rhythms of many units (clonus) or group the corresponding impulses in regular bursts (tremor) The electromyogram is useful in the study of diseases that reduce the number of motor units (poliomyelitis nerve injury) or diminish the number of muscle fibers in each unit (dystrophy myasthenia) or abolish the nerve supply and leave isolated muscle fibers to twitch weakly independently and spontaneously (fibrillation) *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY ELECTRODIAGNOSIS [D D B]

Electron

An elementary particle which is the negatively charged constituent of ordinary matter The electron is the lightest known particle which possesses an electric charge Its rest mass is $m_e \simeq 9\,1 \times 10^{-28}$ g about 1/1836 of the mass of the proton or neutron which are respectively the positively charged and neutral constituents of ordinary matter Discovered in 1895 by Sir J J Thomson in the form of cathode rays the electron was the first elementary particle to be identified *See* ATOMIC STRUCTURE AND SPECTRA CATHODE RAYS CHARGE ELECTRIC ELEMENTARY PARTICLE NUCLEAR STRUCTURE

Charge The charge of the electron is $-e \cong -4\,8 \times 10^{-10}$ esu $-1\,6 \times 10^{-19}$ coulomb The sign of the electron s charge is negative by convention and that of the equally charged proton is positive This is a somewhat unfortunate convention because the flow of electrons in a conductor is thus opposite to the conventional direction of the current *See* CONDUCTION (ELECTRICITY)

The most accurate direct measurement of e is the celebrated oil drop experiment of R A Millikan (1909) in which the charges of droplets of oil in air are measured by finding the electric field which

balances each drop against its weight The weight of each drop is determined by observing its rate of free fall through the air and using Stokes' formula for the viscous drag on a slowly moving sphere The charges thus measured are integral multiples of e For more precise values of e and m_e see ATOMIC CONSTANTS

Electrons and matter Electrons are emitted in radioactivity (as beta rays) and in many other decay processes for instance the ultimate decay prod icts of all mesons are electrons neutrinos and photons the meson s charge being carried away by the electrons (see BETA RAYS MESON RADIOACTIVITY) The electron itself is completely stable according to all available evidence Electrons contribute the bulk to ordinary matter the volume of an atom is nearly all occupied by the cloud of electrons surrounding the nucleus which occupies only about 10^{-13} of the atom s volume The chemical properties of ordinary matter are determined by the electron cloud The electron obeys the Fermi Dirac statistics and for this reason is often called a *fermion* One of the primary attributes of matter impenetrability results from the fact that the electron being a fermion obeys the Pauli exclusion principle the world would be completely different if the lightest charged particle were a boson that is a particle that obeys Bose Einstein statistics See BOSE EINSTEIN STATISTICS EXCLUSION PRINCIPLE FERMI DIRAC STATISTICS

Spin Every elementary particle possesses an intrinsic angular momentum called its spin The spin of the electron has the magnitude ½ $\hbar$ where $\hbar$ is Planck s constant h divided by 2π An electron thus has two spin states spin up and spin down To describe this the nonrelativistic wave function is a two component function that is a vector in a two dimensional spin space the two linearly independent vectors represent the two possible spin states In 1928 P A M Dirac derived the corresponding relativistic wave equation (Dirac equation) Here the electron wave function must have four components correspondingly for a wave of given momentum there are four internal states In addition to the two valued spin coordinate there is an energy coordinate that is for a momentum p the energy can be $\pm\sqrt{(m c^2)^2 + (pc)^2}$ where c is the velocity of light See ELECTRON SPIN QUANTUM THEORY RELATIVISTIC

Positron The negative energy states were at first an embarrassment for they extend downward indefinitely an electron would cascade indefinitely downward in energy radiating photons Electrons fore [illegible] in er [illegible] alrea [illegible] y more elec trons A new process is possible now such a negative energy electron by absorbing energy can go to a positive energy state This leaves behind a hole in the sea of negative energy electrons [illegible]

properties of an electron except that it appears to have a *positive* charge (because it represents a missing negative charge) This particle is the positron first discovered in 1932 by C D Anderson in a cloud chamber study of cosmic radiation See POSITRON

The electron (sometimes called a negatron) and the positron are on an equal footing if one started with a Dirac wave equation for the positron identifying electrons with holes in the negative energy positron states one would get an equivalent theory The apparent dissymmetry inherent in the construction of the hole theory disappears from the results when the total charge energy and momentum of empty space is defined to be zero actually the dissymmetry can be avoided at all stages in the formalism of quantum field theory See QUANTUM FIELD THEORY, SYMMETRY LAWS (PHYSICS)

Magnetic moment The electron has magnetic properties by virtue of its orbital motion about the nucleus of its atom and its rotation (spin) about its own axis The magnetic moment of the electron is predicted by the Dirac equation to be

$$\mu_D = \frac{e\hbar}{2m_e}$$

The actual moment μ differs from μ_D by a small amount (anomalous magnetic moment) due to electromagnetic radiative corrections $\mu = 1.0011$ μ_D This theoretical value calculated using renormalized quantum field theory agrees with the experimental value See QUANTUM ELECTRODYNAMICS

Other information It would be difficult to list all the articles wherein the electron forms an integral part of the discussion The articles listed in this paragraph and in the preceding discussion are intended to be merely a representative sample For information on the vital role played by the electron in technology and engineering see ELECTRONICS and the articles listed therein For additional information see BAND THEORY OF SOLIDS COMPTON EFFECT, ELECTRICITY, ELECTROMAGNETIC RADIATION, ELECTRON CAPTURE, ELECTRON CONFIGURATION, ELECTRON DIFFRACTION, ELECTRON EMISSION ELECTRON MOTION IN VACUUM ELECTRON OPTICS, FREE ELECTRON THEORY OF METALS MAGNETISM, PARTICLE ACCELERATOR QUANTUM MECHANICS QUANTUM THEORY, NONRELATIVISTIC RELATIVISTIC ELECTRODYNAMICS [C J G]

Electron capture

The process in which an atom or ion passing through a material medium either loses or gains one or more orbital electrons In the passage of charged particles (defined here as nuclei having more or less than Z atomic electrons where Z is the atomic number) through matter the capture (and loss) of electrons is an important process in the slowing down of the particles and therefore has a strong influence on their range Thus a neutral hydrogen atom loses only about half as much en

ergy per centimeter as the positively charged proton in passing through matter consisting of light elements

For the ordinary *charged particles* (α particles protons) the capture process is important only at low energies when the particle velocity is of the order of electron velocities in the stopping material and thus is important at the end of the range For fission fragments however which initially have a large excess of positive charges electron capture occurs immediately and continues throughout the slowing-down process This fact causes the energy loss mechanisms at the latter part of the range to be different for fission fragments and protons or α particles (*see* FISSION NUCLEAR) The heavy ions (nuclei of oxygen argon and so forth with all atomic electrons stripped away) now available are intermediate in mass between fission fragments and the light particles and have higher velocity than fission fragments Thus their energy loss is relatively unaffected by electron capture in the early part of the range but in later stages this process has important consequences

The nuclear capture of electrons (*K* capture) occurs by a process quite different from atomic capture and is in fact a consequence of the general β interaction This general interaction includes β decay (the oldest known β transformation and hence the name) β^+ decay (or positron decay) and *K capture* the latter so called because the electron captured by the nucleus is taken from the *K* shell (the shell nearest the nucleus) of atomic electrons (*see* RADIOACTIVITY) The probability of occurrence of electron capture by the nucleus obviously depends on the amount of time the electrons spend at the nucleus that is on the size of the electron wave function at the nuclear center Since to a very good approximation only electrons with zero orbital angular momentum have a wave function that is finite at the center capture is not expected from any b [illegible] K [illegible] ess c [illegible] thus s [illegible] *K* shell) is captured with the simultaneous transition of a *p* electron (from the *L* shell) to the *K* shell with the emission of γ radiation This differs from ordinary x radiation following *K* capture by the fact that energy is not conserved in the transition (it must of course be conserved in the whole process) Since *K* electrons spend more time in the nucleus for large *Z* nuclei than for small *K* capture is more probable in heavier nuclei The second order process (called *L* capture) is even more strongly *Z* dependent and actually controls the shape of the γ ray spectrum for very heavy nuclei

What has been concluded so far depends on the atomic (that is electromagnetic) interactions But the process itself is a result of the β interaction which is between the electron field and the nucleon field (*see* QUANTUM FIELD THEORY) This weak interaction (so called because the processes involving the interaction take place in times that are long on the nuclear time scale) which couples electrons (or positrons) with nuclei γ rays and neutrinos has been the subject of increased study because it has been shown to demonstrate nonconservation of parity The electron and positron are identical except for electromagnetic interactions thus a nucleus which is energetically capable of *K* capture will also usually be capable of positron emission and the two processes do indeed compete in several nuclei *See* AUGER EFFECT PARITY (QUANTUM MECHANICS) [M H H]

Electron configuration

The orbital arrangement of an atom s electrons The electron configuration of an atom specifies the quantum numbers of the atom s electrons in a given energy state Only two types of quantum numbers are needed to describe the electron configuration of any atom or ion One is called the principal or total quantum number n it is an integer and represents the shell in which the electron finds itself Successive shells are numbered 1 2 3 4 5 6 7 and are symbolized K L M N O P Q respectively The other quantum number l is also integral but represents orbital angular momentum of an electron in units of $h/2\pi$ it has values 0 1 2 3 corresponding to s p d f electrons respectively These four l values and the first seven n values suffice to describe the normal electron configurations of all possible atoms and ions that is a total of 5050 for the first 100 chemical elements *See* QUANTUM NUMBERS

In any configuration the number of equivalent electrons (same n and l) is indicated by an integral exponent (not a quantum number) attached to the letters s p d and f according to the Pauli exclu [illegible]

energy state) of sodium means that there are two electrons with $n = 1$ $l = 0$ two with $n - 2$ $l = 0$ six with $n = 2$ $l = 1$ and one with $n = 3$ $l = 0$. Higher energy states result from an increase in any of these quantum numbers

As a consequence of quantum mechanics a limited number of energy states or spectral terms are naturally associated with each electron configuration and the latter are uniquely and unambiguously disclosed and designated by certain properties (frequencies and intensities) of the radiations and by certain features (multiplicities intervals magnetic splitting factors) of the spectral terms *See* QUANTUM MECHANICS

An electron configuration is categorized as even or odd according to whether the sum of p and f electrons is even or odd and spectral lines result only from transitions between configurations of unlike parity *See* PARITY (QUANTUM MECHANICS)

Insofar as they are known from spectroscopic investigations the electron configurations characteristic of the normal or ground states of 100 chemical elements are shown in the accompanying table In the next to last column of the table the

Distribution of electrons in the atoms

Element		K	L		M			N				O				Ground term	Ionization potential ev
		1 0 1s	2 0 2s	2 1 2p	3 0 3s	3 1 3p	3 2 3d	4 0 4s	4 1 4p	4 2 4d	4 3 4f	5 0 5s	5 1 5p	5 2 5d	5 3 5f		
H	1	1		–		–					–	–	–	–	–	$^2S_{1/2}$	13.595
He	2	2														$^1S_0$	24.58
Li	3	2	1		–	–	–				–	–	–	–	–	$^2S_{1/2}$	5.390
Be	4	2	2										–	–	–	$^1S_0$	9.320
B	5	2	2	1								–	–	–	–	$^2P^o_{1/2}$	8.296
C	6	2	2	2			–					–	–	–		$^3P_0$	11.256
N	7	2	2	3									–	–		$^4S^o_{3/2}$	14.53
O	8	2	2	4							–		–	–	–	$^3P_2$	13.614
F	9	2	2	5	–								–	–	–	$^2P^o_{3/2}$	17.418
Ne	10	2	2	6										–	–	$^1S_0$	21.559
Na	11				1				–	–		–		–	–	$^2S_{1/2}$	5.138
Mg	12				2								–	–	–	$^1S_0$	7.644
Al	13				2	1							–	–	–	$^2P^o_{1/2}$	5.984
Si	14		Neon configuration		2	2					–		–	–	–	$^3P_0$	8.149
P	15				2	3						–	–	–	–	$^4S^o_{3/2}$	10.484
S	16				2	4							–	–	–	$^3P_2$	10.357
Cl	17				2	5	–			–			–	–	–	$^2P^o_{3/2}$	13.01
Ar	18				2	6	–			–			–	–	–	$^1S_0$	15.755
K	19							1				–	–	–		$^2S_{1/2}$	4.339
Ca	20							2				–	–	–		$^1S_0$	6.111
Sc	21						1	2					–	–	–	$^2D_{3/2}$	6.54
Ti	22						2	2				–	–	–	–	$^3F_2$	6.82
V	23						3	2					–		–	$^4F_{3/2}$	6.74
Cr	24						5	1					–		–	$^7S_3$	6.764
Mn	25						5	2					–		–	$^6S_{5/2}$	7.432
Fe	26						6	2		–		–	–		–	$^5D_4$	7.87
Co	27		Argon configuration				7	2					–		–	$^4F_{9/2}$	7.86
Ni	28						8	2					–		–	$^3F_4$	7.633
Cu	29						10	1					–		–	$^2S_{1/2}$	7.724
Zn	30						10	2					–		–	$^1S_0$	9.391
Ga	31						10	2	1			–	–	–		$^2P^o_{1/2}$	6.00
Ge	32						10	2	2				–	–		$^3P_0$	7.88
As	33						10	2	3				–	–	–	$^4S^o_{3/2}$	9.81
Se	34						10	2	4				–		–	$^3P_2$	9.75
Br	35						10	2	5				–		–	$^2P^o_{3/2}$	11.84
Kr	36						10	2	6	–	–		–			$^1S_0$	13.996
Rb	37											1	–		–	$^2S_{1/2}$	4.176
Sr	38									–		2	–		–	$^1S_0$	5.692
Y	39									1		2	–	–		$^2D_{3/2}$	6.38
Zr	40									2		2	–		–	$^3F_2$	6.84
Nb	41		Krypton configuration							4	–	1	–		–	$^6D_{1/2}$	6.88
Mo	42									5		1	–	–	–	$^7S_3$	7.10
Tc	43									5		2	–	–	–	$^6S_{5/2}$	7.28
Ru	44									7		1	–	–	–	$^5F_5$	7.364
Rh	45									8	–	1	–	–	–	$^4F_{9/2}$	7.46
Pd	46									10		–	–	–	–	$^1S_0$	8.33

spectral term of the energy level with zero energy (normal unexcited state of the atom) is shown. The main part of the term symbol is a capital letter S, P, D, F, etc. (representing the resultant l value) to which is attached a superior prefix 1, 2, 3, 4, etc. (indicating the multiplicity) and an anterior suffix 0, ½, 1, 3/2, 2, 5/2 (showing the total angular momentum or J value of the atom in this particular state). A sign ^{o} above the J value signifies that the spectral term and electron configuration have odd parity.

The last column of the table presents the first ionization potential of the atom when this has been derived from spectroscopic observations. In any atomic spectrum two or more spectral lines with certain similar properties may form a series such that the reciprocal wavelengths $1/\lambda$ (number of waves per cm = σ) can be closely represented by a formula of the Rydberg type, $\sigma = L - R/(n+\mu)^2$, in which L is the limit of the series, R is called the Rydberg constant and n (the principal quantum number) has successive integral values to which a

Distribution of electrons in the atoms (Cont.)

Element		Configuration of inner shells	N	O				P			Q	Ground term	Ionization potential ev
			4 3 4f	5 0 5s	5 1 5p	5 2 5d	5 3 5f	6 0 6s	6 1 6p	6 2 6d	7 0 7s		
Ag	47	Palladium configuration	–	1	–				–			$^2S_{1/2}$	7 574
Cd	48		–	2								$^1S_0$	8 991
In	49		–	2	1							$^2P^\circ_{1/2}$	5 785
Sn	50		–	2	2			–	–			$^3P_0$	7 342
Sb	51		–	2	3						–	$^4S^\circ_{3/2}$	8 639
Te	52		–	2	4							$^3P_2$	9 01
I	53		–	2	5							$^2P^\circ_{3/2}$	10 453
Xe	54		–	2	6							$^1S_0$	12 127
Cs	55	The shells 1s to 4d contain 46 electrons	–	The shells 5s to 5p contain 8 electrons				1			–	$^2S_{1/2}$	3 893
Ba	56							2				$^1S_0$	5 210
La	57		–			1		2			–	$^2D_{3/2}$	5 61
Ce	58		(2)					(2)					
Pr	59		3					2				$^4I^\circ_{9/2}$	
Nd	60		4					2				$^5I_4$	
Pm	61		(5)					(2)					
Sm	62		6			–		2			–	$^7F_0$	5 6
Eu	63		7					2				$^8S^\circ_{7/2}$	5 67
Gd	64		7			1		2			–	$^9D^\circ_2$	6 16
Tb	65		(9)					(2)	–				
Dy	66		(10)					(2)		–			
Ho	67		(11)					(2)			–		
Er	68		(12)					(2)					
Tm	69		13					2		–		$^2F^\circ_{7/2}$	
Yb	70		14					2			–	S_0	6 2
Lu	71		14			–	–	2				$^2D_{5/2}$	6 15
Hf	72	The shells 1s to 5p contain 68 electrons				2		2			–	$^3F_2$	6 8
Ta	73					3		2				$^4F_{3/2}$	7 88
W	74					4		2		–		$^5D_0$	7 98
Re	75					5		2	–	–		$^6S_{5/2}$	7 87
Os	76					6		2			–	$^5D_4$	8 7
Ir	77					7		2				$^4F_{9/2}$	9 0
Pt	78					9	–	1		–		$^3D_3$	9 0
Au	79	The shells 1s to 5d contain 78 electrons						1	–			$^2S_{1/2}$	9 22
Hg	80							2			–	$^1S_0$	10 43
Tl	81							2	1		–	$^2P^\circ_{1/2}$	6 106
Pb	82						–	2	2			$^3P_0$	7 415
Bi	83						–	2	3			$^4S^\circ_{3/2}$	7 287
Po	84						–	2	4			$^3P_2$	8 43
At	85							(2)	(5)				
Rn	86							2	6			$^1S_0$	10 746
Fr	87							2	6		(1)		
Ra	88							2	6		2	$^1S_0$	5 277
Ac	89							2	6	1	2	$^2D_{3/2}$	6 9
Th	90							2	6	2	2	$^3F_2$	
Pa	91						(2)	2	6	(1)	(2)		
U	92						3	2	6	1	2	$^5L_6$	~4 0
Np	93						(4)	2	6	1	(2)		
Pu	94						(6)	2	6		2	$^7F_0$	
Am	95						7	2	6	–	2	$^8S^\circ_{7/2}$	6 0
Cm	96						7	2	6	1	2	$^9D^\circ_2$	
Bk	97						(8)	2	6	(1)	2		
Cf	98						(9)	2	6	(1)	2		
Es	99						(10)	2	6	(1)	2		
Fm	100						(11)	2	6	(1)	2		

constant fractional part μ is added (*see* RYDBERG CONSTANT). The second term vanishes when n approaches infinity and the series limit is thus evaluated. This limit is usually coincident with the ground state of the ion and is thus a measure (in wave number units) of the energy required to remove from an atom its least firmly bound electron and transform a neutral atom into a singly charged ion. The energy required to ionize an atom is usually expressed in electron volts (1 ev = 8066 wave

numbers) and is called its first ionization potential See IONIZATION POTENTIAL see also ATOMIC STRUCTURE AND SPECTRA [E A J W F M]

Electron diffraction

The phenomenon associated with the interference processes which occur when electrons are scattered by atoms to form diffraction patterns The wave character of electrons is shown most strikingly and doubtless most conclusively by the phenomena of interference For this reason the diffraction of electrons presents the most obvious confirmation of the new quantum mechanics This is probably the most important result that has come from the observation that electrons are diffracted by crystals A second aspect of electron diffraction is its use as a research tool in studying the structure of crystals and of free molecules somewhat analogous to the use of x rays for these purposes See X RAY CRYSTALLOGRAPHY X RAY DIFFRACTION

Theory According to quantum theory any particle moving with momentum mv has a wavelength $\lambda = h/mv$ where h is Planck's constant (see DE BROGLIE WAVELENGTH) If the particle is an electron and its velocity is the result of having fallen through the potential difference V this formula becomes

$$\lambda = \frac{h}{(2m_0eV)^{1/2}(1 + eV/2m_0c^2)^{1/2}}$$

where m_0 and e are the rest mass and charge of the electron and c is the velocity of light For V in volts and λ in angstroms A

$$\lambda = \frac{12.2638}{V^{1/2}(1 + 9.7852 \times 10^{-7}V)^{1/2}} \qquad (1)$$

The last factor in the denominators of these equations represents the relativity correction which is negligible at low voltages and amounts to only 5% at 100 000 volts

This formula can be tested by measurements upon a diffraction pattern such as that shown in Fig 1 which was made by a beam of 50-kv electrons scattered in passing through a polycrystalline foil of a gold copper alloy about 400 A thick One compares the wavelength obtained from the formula with V equal to 50 000 volts with that calculated from the Bragg equation $\lambda = 2d_{hkl} \sin \theta_{hkl}$ where the values of d_{hkl} are the known spacings between adjacent lattice planes and the values of θ_{hkl} are diffraction angles corresponding to the rings of the pattern The best experiments of this type have been carried out with sufficient care to check the wavelength given by quantum theory to a precision of about 0.1% See QUANTUM MECHANICS QUANTUM THEORY, NONRELATIVISTIC

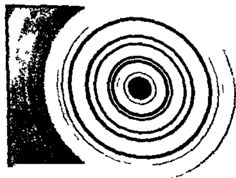


Fig 1 Electron-diffraction pattern of a gold-copper alloy

Experimental techniques. In the use of electron diffraction as a research tool there are two widely separated fields of application One is the study of the structures of the surfaces of solid bodies and the other the investigation of the interatomic spacings in gaseous and liquid molecules

DIFFRACTION IN SOLIDS

The voltages used for electron diffraction in solids vary from less than 40 to over 100 000 corresponding to a wavelength range from greater than 2 A to less than 0.04 A The experimental techniques best suited for electrons are so different at low and high voltages that this range is divided arbitrarily into two parts the low voltage range (~ 100 volts or 1 A) and the high voltage range ... elec-

ments to determine a diffraction pattern At ... voltages a pattern is recorded directly on a photographic plate as in the pattern of Fig 1 The photographic technique is much simpler and this has led to attempts to accelerate the electrons of an entire low voltage diffraction pattern until they darken a photographic plate This procedure which could facilitate experimentation with slow electrons seems amenable to practical development

Refraction effects In all applications consideration must be given to the refraction of electron waves Within a solid body the electron wavelength λ' is obtained from Eq (1) by replacing V by $(V + \Phi)$ where Φ, the inner potential is of the order of 10 volts The refractive index with respect to empty space is $\mu = \lambda/\lambda'$ (a form of Snell's law) which is equal to $(1 + \Phi/V)^{1/2}$ to a high degree of precision For low voltages μ is quite large and the interpretation of diffraction patterns produced by low voltage electrons must take into account the resulting large displacements of individual diffraction beams

There are moreover remarkable anomalies associated with the fact that a single average value of the inner potential is not realistic when only two or three layers of atoms are considered For high velocity electrons μ is close to unity and in many cases no detectable displacement of diffraction beams results as in transmission through thin films In the case of scattering of electrons from the surface of a massive body however the glancing angle must be so small that relatively large refractive effects occur

Applications Because electrons are scattered very effectively by matter, their penetration is slight Experimentation must be carried out in vacuum and it can give information regarding the [illegible] vestigated by any other diffraction technique The upper range of thickness which can be studied varies greatly with electron speed from about 500 A for 50-kv electrons and heavy metals down to two or three layers of atoms for 100-volt electrons The latter figure is so low that low speed electrons can be profitably used only in very high vacuum and the technique is used only for research It has been less extensively exploited than has the diffraction of fast electrons but its use will probably increase because of widespread interest in the surface states of semiconductors The lower limit of sensitivity is very low indeed, for 50 kv electrons it corresponds to a mean surface thickness for nucleated material not greater than 0 5 A

Studies of structures by means of electrons of potentials of the order of 50 kv are carried out in diffraction cameras made for the purpose and in standard electron microscopes which are in general admirably adaptable (*see* MICROSCOPE ELECTRON) With the latter equipment an electron photomicrograph can be made of the same specimen The diameter of the electron beam on the specimen can be made less than 1 micron This permits exploration of a specimen which is made up of two or more segregated phases in order to obtain photo micrographs and diffraction patterns of the separate phases a technique which is known as selected area diffraction An example of a diffraction pattern showing two separate phases is reproduced as Fig 2 This was made by the reflection method from the surface of a single crystal of Alnico V a magnetic alloy The larger widely spaced diffraction spots were produced by the iron rich matrix and the smaller closely spaced spots by a cobalt rich precipitate of face-centered cubic symmetry having a unit cell dimension $a_0 = 10.0$ A The alloy owes its high coercive force to this precipitate

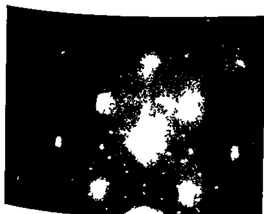


Fig 2 Diffraction pattern of Alnico V

Failure to find the precipitate by x ray analysis is probably due to the smallness of the individual crystals

In many applications electron diffraction owes its value more directly to the slight penetrating power of electrons than is the case in the illustration just given A few examples can be cited One can study the oxidation and corrosion of metal surfaces The beginning of crystal growth can be studied and the beginning of the ordering process in a disordered alloy ordering is found to occur more rapidly and at much lower temperatures in thin films than in massive specimens Orientation of single layers of organic molecules on metal surfaces can be investigated as can the effect of friction on such layers Some partially degenerate crystal structures have been studied in particular carbon containing considerable hydrogen Nucleation of vaporized films has been studied, as has the ability of metal atoms to move over surfaces to form three dimensional crystals Protective layers of aluminum hydrate have been found to form on a silica surface affording some protection from silicosis in the case of ingested silica particles An important development is high resolution diffraction in which fine structure is found within a single diffraction beam Structure of this sort has given information about stacking faults in crystals strains due to dislocations in metals and many other phenomena [L H G]

DIFFRACTION IN GASES AND LIQUIDS

Electron diffraction in gases and liquids is similar in principle to that in solids, the differences arise from the lack in gases and liquids of any highly regular arrangement of the component atoms In gases the low density makes it possible to study diffraction by individual atoms and molecules The results obtained from monatomic gases represent the density of electronic charge in the atom as a function of the distance from the nucleus The results from gaseous polyatomic molecules represent the equilibrium distances between the atomic nuclei and the average amplitudes of vibration associated with these distances Liquids have been studied much less thoroughly, both in theory and in practice than have gases

Applications Typical questions of molecular structure studied by electron diffraction include those of configuration and size in many molecules with special interest in the variation of chemical bond distances in related molecules (such as the 0 068 A decrease in C—F distance in the series CH_3F CH_2F_2, CHF_3, CF_4), the variation in the angles between chemical bonds the distinction between geometric isomers the degree of restricted rotation around chemical bonds and in general the relations between the geometry of molecules and their energy and chemical behavior

The application of electron diffraction in liquids has been restricted because of the less precise information obtainable from molecules which lie in close contact but in irregular arrangements and

because of the limitation to liquids that have low

c
n
e

separation between more distant neighbors varies considerably Evaluation of the closest distance of approach can be made to about 0 1 A major developments in the theory are still required

The most interesting application of electron diffraction to a liquid or amorphous condensed phase has been in the examination of the surface layers on polished solids The diffuse diffraction patterns often obtained may be the result of an amorphous arrangement in the surface or of the poor resolving power of very tiny crystalline particles It is probable that both these states are produced in the polishing of different solids

Theory and techniques Structural information about gaseous molecules is obtained by having a fine beam of electrons pass through the gas and strike a photographic plate The electrons in the beam interact with the charged particles in the atoms (electrons and nuclei) and are bent away from the original direction through varying angles as registered in the pattern on the plate The observed variation in the number of scattered electrons with increasing angle is interpreted as an interference effect that is electron scattering can be described in the language of diffracted waves for which the resultant is the sum of the component wavelets whose amplitudes and phases are influenced by the wavelength of the incident radiation and the relative positions of the scattering centers The equivalent wavelength of the electrons λ is determined by their energy and is equal to

$$\lambda = h/mv \approx (150/V)^{1/2} \times 10^{-8}\ \text{cm}$$

where m v and V are the mass of the electron the velocity and the accelerating voltage respectively and h is Planck s constant The values of λ commonly used in diffraction by gases are in the range $0\ 07\text{–}0\ 05 \times 10^{-8}$ cm

The intensity I of electrons scattered by a spherically symmetrical atom is computed by the Schrodinger wave equation the result is

$$I = k(Z - F)^2/s^4 = kf \qquad (2)$$

where k is a constant Z is the atomic number (charge on the nucleus)

$$F = 4\pi \int_0^\infty r^2\rho(r)[(\sin sr)/sr]\, dr$$

$\rho(r)$ is density of electronic charge at distance r from the nucleus $s = 4\pi(\sin\theta)/\lambda$ with 2θ equal to the angle between the scattered electron and the original beam λ the electron wavelength and f is called the atomic scattering factor Experimental observations on I as a function of s in monatomic gases lead to the determination of $\rho(r)$

The intensity of electrons scattered by a collection of independent molecules having all possible orientations is given by

$$I(s) = k\{\sum_i f_i^2 + \sum_i \sum_j f_i f_j \int_0^\infty P_{ij}(r)[(\sin sr)/sr]\, d\} \qquad (3)$$

where the summations are taken over all the atoms in the molecule The double summation has a term for each pair of atoms i and j The relative probability of finding the distance between the atom i and j at various values is represented by $P_{ij}(r)$ In principle the use of the observed intensity to de er mine P_{ij} for each pair of atoms in the molecule constitutes a structure determination for the molecule The expression is simpler when the atom c cule The expression is simpler when the atomic motions are harmonic as in the case of CCl_4 For this molecule the double summation has only two distinct terms The first one is

$$8\, f_C f_C \exp(-l_{CCl} s^2)\ (\sin s r_{CCl})/s r_{CCl}$$

The four
are the
e aver
Molec
ed with

ular parameters such as these can be with high precision when the necessary corrections to the formulas are made for the effect of incoherent scattering and the effect of any large d fferences in atomic numbers among the nuclei in one molecule

Special equipment is required to meet the conditions assumed in the scattering theory The electron beam is accelerated with a steady voltage between 30 000 and 70 000 volts which should be constant within about 0 01% The beam is focused by electrostatic and magnetic lenses o that it h a diameter of no more than 0 1 mm at the photographic plate The whole path of the beam is e closed in a high vacuum chamber (pressure about 10^{-5} mm of mercury) so that no appreciable electron scattering will occur in the residual a The gas specimen is introduced in a fine jet so th the volume in which the electrons meet the gas no more than 0 2 mm in diameter high sp pumping is required to remove the gas as rapi as possible In front of the photographic plat rotating sector is mounted to modify the inten of electrons reaching the plate so that the r mally rapid decrease of intensity with increa angle of scattering is leveled off in a known angle of scattering is able to register the incid and the emulsion is able to register the incid electrons over a wide range of angle

Interpretation of results The pattern obse is a set of light and dark circular bands whose s ings and relative intensities depend on the com tion and structure of the specimen molecules. pattern is scanned by a recording microphoton which yields (after calibration of the photogra emulsion) a tracing of the experimental da I as a function of s These data are interp with the aid of Eq (3) to give the P_{ij} func for the pairs of atoms as illustrated for CC Fig 3 The two prominent peaks represen C—Cl and Cl—Cl distances the small devi from the background line are caused by erro the experimental data and in the method of

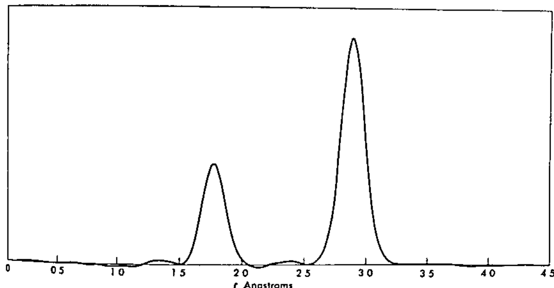


Fig 3 Experimental distribution of internuclear distances in CCl_4

pretation From the positions and widths of the peaks it is determined that

$$r_{CCl} = 1\ 766 \pm 0\ 003\ A$$
$$l_{CCl} = 0\ 060 \pm 0\ 005\ A$$
$$r_{ClCl} = 2\ 887 \pm 0\ 004\ A$$
$$l_{ClCl} = 0\ 068 \pm 0\ 003\ A$$

This relatively high precision in determining the distances in gas molecules has been achieved since about 1950 with the development of new instrumentation and the application of high speed computing methods The intensities of scattered electrons registered on the photographic plate can now be measured to about 1%, the time saved by the use of punched card computing methods in interpreting the data permits a more rigid application of the criteria for satisfactory agreement between experiment and theory Through 1958 about 40 gases had been studied by the methods giving internuclear distances within 0 003 A, when these substances have also been investigated by spectroscopic methods good agreement is found between the results of the two methods if allowance is made for the difference in the nature of the distances measured The molecular structures of more than 500 gases have been reported by earlier electron diffraction procedures with uncertainties 10 times larger or more

The applicability of the gas diffraction method alone is limited to simple molecules Only three distinct distances in a molecule can be determined with high precision, light atoms in the presence of heavy ones in the same molecule are less precisely located, distances as close as 0 03 A can barely be resolved The range of the method is greatly increased, however, when some structural features in the molecule can be assumed from the results of other methods of investigation For example, if a 6-membered ring of atoms is known to have trigonal symmetry the number of structural parameters is decreased from 12 to 2 *See* NEUTRON DIFFRACTION, SCATTERING EXPERIMENTS, ATOMIC AND MOLECULAR, SCATTERING EXPERIMENTS NUCLEAR

[L O B]

Bibliography Z G Pinsker, *Electron Diffraction* 1953, L E Sutton (ed), *Interatomic Distances* 1958, G P Thomson and W Cochrane, *Theory and Practice of Electron Diffraction*, 1939, A Weissberger (ed) *Physical Methods of Organic Chemistry*, 3d ed 1959

Electron emission

The liberation of electrons from a substance into vacuum Since all substances are built up of atoms and since all atoms contain electrons, any substance may emit electrons, usually, however, the term refers to emission of electrons from the surface of a solid

The process of electron emission is analogous to that of ionization of a free atom, in which the latter parts with one or more electrons The energy of the electrons in an atom is lower than that of an electron at rest in vacuum, consequently, in order to ionize an atom energy must be supplied to the electrons in some way or other By the same token, a substance does not emit electrons spontaneously, but only if some of the electrons have energies equal to or larger than that of an electron at rest in vacuum This may be achieved by various means If a substance is heated, the atoms begin to vibrate with larger amplitudes, and electrons may absorb

emission from a substance may be induced by bombardment with charged particles such as electrons or ions in the phenomenon called secondary emission Field emission, or cold emission, refers to the emission of electrons under influence of a

because of the limitation to liquids that have low

molecules in the liquid state the distances between nearest neighbors are fairly uniform whereas the separation between more distant neighbors varies considerably Evaluation of the closest distance of approach can be made to about 0 1 A major developments in the theory are still required

The most interesting application of electron diffraction to a liquid or amorphous condensed phase has been in the examination of the surface layers on polished solids The diffuse diffraction patterns often obtained may be the result of an amorphous arrangement in the surface or of the poor resolving power of very tiny crystalline particles It is probable that both these states are produced in the polishing of different solids

Theory and techniques Structural information about gaseous molecules is obtained by having a fine beam of electrons pass through the gas and strike a photographic plate The electrons in the beam interact with the charged particles in the atoms (electrons and nuclei) and are bent away from the original direction through varying angles as registered in the pattern on the plate The observed variation in the number of scattered electrons with increasing angle is interpreted as an interference effect that is electron scattering can be described in the language of diffracted waves for which the resultant is the sum of the component wavelets whose amplitudes and phases are influenced by the wavelength of the incident radiation and the relative positions of the scattering centers The equivalent wavelength of the electrons λ is determined by their energy and is equal to

$$\lambda = h/mv \approx (150/V)^{1/2} \times 10^{-8} \text{ cm}$$

where m v and V are the mass of the electron the velocity and the accelerating voltage respectively and h is Planck s constant The values of λ commonly used in diffraction by gases are in the range 0 07 0 05 $\times 10^{-8}$ cm

The intensity I of electrons scattered by a spherically symmetrical atom is computed by the Schrodinger wave equation the result is

$$I = k(Z - F)^2/s^4 = kf^2 \qquad (2)$$

where k is a constant Z is the atomic number (charge on the nucleus)

$$F = 4\pi \int_0^\infty r^2\rho(r)\,[(\sin sr)/sr]\,dr$$

$\rho(r)$ is density of electronic charge at distance r from the nucleus $s = 4\pi(\sin\theta)/\lambda$ with 2θ equal to the angle between the scattered electron and the original beam λ the electron wavelength and f is called the atomic scattering factor Experimental observations on I as a function of s in monatomic gases lead to the determination of $\rho(r)$

The intensity of electrons scattered by a collection of independent molecules having all possible orientations is given by

$$I(s) = k\{\sum_i f_i^2 + \sum_i \sum_j f_i f_j \int_0^\infty P_{ij}(r)\,[(\sin sr)/sr]\,d\} \qquad (3)$$

where the summations are taken over all the atoms in the molecule The double summation has a term for each pair of atoms i and j The relative probability of finding the distance between the atom i and j at various values is represented by $P_{ij}(r)$ In principle the use of the observed intensity to determine P_{ij} for each pair of atoms in the molecule constitutes a structure determination for the molecule The expression is simpler when the atomic

$$8\, f_C f_{Cl} \exp(-l_{CCl}\, s^2)\ (\sin s r_{CCl})/s r_{CCl}$$

and the second one for Cl—Cl is similar The four parameters which describe the structure are the equilibrium distances r_{CCl} and r_{ClCl} and the average amplitudes of vibration l_{CCl} and l_{ClCl} Molecular parameters such as these can be reported with high precision when the necessary corrections to the formulas are made for the effect of incoherent scattering and the effect of any large differences in atomic numbers among the nuclei in one molecule

Special equipment is required to meet the conditions assumed in the scattering theory The electron beam is accelerated with a steady voltage between 30 000 and 70 000 volts which should be constant within about 0 01% The beam is focused by electrostatic and magnetic lenses so that it has a diameter of no more than 0 1 mm at the photographic plate The whole path of the beam is enclosed in a high vacuum chamber (pressure of about 10^{-5} mm of mercury) so that no appreciable electron scattering will occur in the residual air The gas specimen is introduced in a fine jet so that the volume in which the electrons meet the gas is no more than 0 2 mm in diameter high-speed pumping is required to remove the gas as rapidly as possible In front of the photographic plate a rotating sector is mounted to modify the intensity of electrons reaching the plate so that the normally rapid decrease of intensity with increasing angle of scattering is leveled off in a known way and the emulsion is able to register the incident electrons over a wide range of angle

Interpretation of results The pattern observed is a set of light and dark circular bands whose spacings and relative intensities depend on the composition and structure of the specimen molecules The pattern is scanned by a recording microphotometer which yields (after calibration of the photographic emulsion) a tracing of the experimental data of I as a function of s These data are interpreted with the aid of Eq (3) to give the P_{ij} functions for the pairs of atoms as illustrated for CCl_4 in Fig 3 The two prominent peaks represent the C—Cl and Cl—Cl distances the small deviations from the background line are caused by errors in the experimental data and in the method of inter

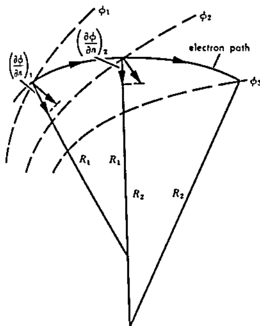


Fig 1 Path plotting in an electrostatic field The dashed lines ϕ_1 ϕ_2 and ϕ_3 are equipotential lines R_1 and R_2 are radii of curvature of the electron path

lytically by solving Laplace s equation

$$\nabla^2\phi = \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} = 0$$

More generally it can be found by constructing a large scale model of the electrode structure and immersing it in an electrolytic tank so that the surface of the liquid (usually slightly acidified tap water) coincides with the plane of symmetry of interest With potentials proportional to the actual potentials applied to the model electrodes an equipotential line on the surface can be found by determining the points at which a probe at the potential in question draws no current

The path equation can be shown to be equivalent to the expression for the radius of curvature R of the paths

$$R = \frac{2\phi}{(\partial\phi/\partial n)}$$

where $-\partial\phi/\partial n$ is the component of the electric field normal to the electron path If an equipotential plot has been prepared this relation permits the graphical plotting of an electron path (Fig 1) The path is approximated by a series of circular arcs the radius of curvature between successive equipotential lines being computed from the preceding relation for R

Paraxial ray equation Electrostatic fields which have not only a plane of symmetry but an axis of symmetry which represents the intersection of an infinite family of planes of symmetry have particular practical importance The path equation previously given still applies here provided that y is identified with r the distance from the axis and x with z the distance measured along the axis The Laplace equation in the new coordinates z r takes the form

$$\frac{\partial^2\phi}{\partial r^2} + \frac{1}{r}\frac{\partial\phi}{\partial r} + \frac{\partial^2\phi}{\partial z^2} = 0$$

This equation is solved quite generally by the series

$$\phi(z\ r) = \Phi(z) - \frac{r^2}{4}\frac{d^2\Phi}{dz^2} + \frac{r^4}{64}\frac{d^4\Phi}{dz^4}$$

Here $\Phi(z)$ is the potential on the axis of symmetry Thus the potential everywhere within the axially symmetric electrode structure is fully determined by the potential variation along the axis Substitution of the expansion of ϕ in the path equation with retention of terms of the first order only in r and dr/dt leads to the paraxial ray equation

$$\frac{d^2r}{dz^2} + \frac{1}{2}\frac{d\Phi}{dz}\frac{dr}{dz} + \frac{1}{4}\frac{d^2\Phi}{dz^2}r = 0$$

This is the path equation for electrons whose paths depart relatively little both in slope and in distance from the axis of the field

The paraxial equation is linear in r This means that if one electron path intersects the axis in two points all electron paths passing through one of the points also pass through the other In brief the

tric field acts on the paths of electrons in the same manner as glass lenses act on light rays (*see* ELECTROSTATIC LENS) Departures of the exact path from the paraxial equation result in image defects or aberrations similar in character to those observed for glass lenses

Magnetic fields A magnetic field exerts on an electron of velocity **v** a force **F** which is perpendicular to both the direction of motion and the direction of the field In vector notation this Lorentz force is given by

$$\mathbf{F} = -e(\mathbf{v} \times \mathbf{b})$$

Here **b** is the magnetic induction The components of the Lorentz force are

$$F_x = e\left[b_y\frac{dz}{dt} - b_z\frac{dy}{dt}\right]$$

$$F_y = e\left[b_z\frac{dx}{dt} - b_x\frac{dz}{dt}\right]$$

$$F_z = e\left[b_x\frac{dy}{dt} - b_y\frac{dx}{dt}\right]$$

Since this force is perpendicular to the direction of motion it does no work on the electron whose velocity consequently remains unchanged in magnitude

For a uniform magnetic field parallel to the z axis

$$b_z = B \qquad b_x = b_y = 0$$

Newton's second law leads to a constant z component of the velocity The magnitude of the velocity component v_{xy} in the xy plane is similarly constant since the square of the total velocity is equal to the sum of the squares of the components For the motion projected on the xy plane Newton s second law thus takes the form

$$mv_{xy}^2/R = ev_{xy}B$$

Here R is the radius of curvature of the projected path R is seen to be a constant so that the projected path is a circle with radius

$$R = \frac{mv_{xy}}{eB} = \frac{\sin\alpha}{B}\left(\frac{2m\phi}{e}\right)^{1/2} = \frac{3\ 37\phi^{1/2}}{B}\sin\alpha \quad \text{cm}$$

Here α is the angle which the electron path makes with the field direction and ϕ is the accelerating potential of the electrons B is measured in gauss and ϕ in volts The frequency with which the circle is traversed by the electron is given by

$$f = v_{xy}/(2\pi R) = eB/(2\pi m) = 2\,8 \times 10^6 B \quad \text{sec}^{-1}$$

This frequency the cyclotron frequency thus depends only on the magnetic field strength

The complete motion of the electron (Fig 2) is thus a helix about a magnetic line of force with a pitch

$$d = v_z/f = \frac{2\pi\cos\alpha}{B}\left(\frac{2m\phi}{e}\right)^{1/2} = 21\ 08\frac{\phi^{1/2}}{B}\cos\alpha \quad \text{cm}$$

All electrons passing through a point with equal axial velocity components pass through a series of points separated by d on the same magnetic field line An initially divergent electron beam is held together by a uniform magnetic field since an electron path which intersects a particular field line can never depart from it by more than twice the radius R of the helix Uniform magnetic fields are widely used for keeping electron beams from spread [illegible]
klys [illegible]
field [illegible]
is u [illegible] magnetic deflection of beams in certain television camera tubes such as the image orthicon and the vidicon In these tubes a weak transverse magnetic deflection field is superposed on a strong longitudinal magnetic focusing field *See* TELEVISION CAMERA TUBE

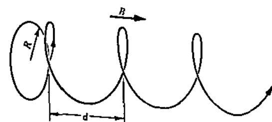


Fig 2 Motion of an electron in a uniform magnetic field Its path is in general a helix with pitch d, radius R and axis parallel to the field

Motion in nonuniform magnetic fields with axial symmetry is conveniently treated as a special case of motion in combined electric and magnetic field

Combined fields with axial symmetry The equations of motion now are

$$m\frac{d^2z}{dt^2} = e\left[\frac{\partial\phi}{\partial z} + b_r r\frac{d\theta}{dt}\right]$$

$$m\left[\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2\right] = e\left[\frac{\partial\phi}{\partial r} - b_z r\frac{d\theta}{dt}\right]$$

$$m\frac{1}{r}\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right) = e\left[b_z\frac{dr}{dt} - b_r\frac{dz}{dt}\right]$$

The coordinates z, r, and θ represent distance along the axis perpendicular distance from the axis and azimuthal angle about the axis The terms b_z and b_r are the axial and radial components of the magnetic induction From these equations a path equation expressing the variation of the radial distance r with the axial distance z is derived One obtains

$$\frac{d^2r}{dz^2} = \frac{1}{2\phi^*}\left[1 + \left(\frac{dr}{dz}\right)^2\right]\left[\frac{\partial\phi^*}{\partial r} - \frac{dr}{dz}\frac{\partial\phi^*}{\partial z}\right]$$

with $\phi^* = \phi(1 - D^2) = \phi - \left[\frac{C}{r} + \left(\frac{e}{2m}\right)^{1/2} A\right]^2$

where ϕD^2 is a shorthand symbol for the last term in the equation, and C is, except for a universal multiplying constant the angular momentum of the electron about the axis at a point where the magnetic field vanishes

$$C = r^2\frac{d\theta}{dz}\phi^{1/2}\left[\left(\frac{dr}{dz}\right)^2 + r^2\left(\frac{d\theta}{dz}\right)^2 + 1\right]^{-1/2} - \left(\frac{e}{2m}\right)^{1/2} rA$$

Here A is the magnetic vector potential which is numerically equal to the magnetic flux through a circle about the axis through the reference point divided by the circumference of that circle

At the same time, the azimuth θ of the electron changes according to the expression

$$\theta = \theta_0 + \int_{z_0}^{z}\frac{D}{r(1 - D^2)^{1/2}}\left[1 + \left(\frac{dr}{dz}\right)^2\right]^{1/2} dz$$

The path equation can be solved by the graphical and numerical methods useful for determining electron paths in electrostatic fields The general paraxial equation is obtained by substituting the expansion

$$A = \frac{r}{2}B(z) - \frac{r^3}{16}\frac{d^2B(z)}{dz^2}$$

Here $B(z)$ is the magnetic induction along the axis Substitution of this expansion and that for the electrostatic potential and retention of terms

of the first order in r and dr/dz only lead to

$$\frac{d^2r}{dz^2}+\frac{1}{2\Phi}\frac{d\Phi}{dz}\frac{dr}{dz}+\left(\frac{1}{4\Phi}\frac{d^2\Phi}{dz^2}+\frac{eB^2}{8m\Phi}-\frac{C^2}{\Phi r^4}\right)r=0$$

$$\theta-\theta_0+\int_{z_0}^{z}\left[\frac{C}{r^2\Phi^{1/2}}+\left(\frac{e}{8m\Phi}\right)^{1/2}B\right]dz$$

$$\frac{C}{r_0^2\Phi_0^{1/2}}=\left(\frac{d\theta}{dz}\right)_0-\left(\frac{e}{8m\Phi_0}\right)^{1/2}B(z_0)$$

With B in gauss, Φ in volts and z in centimeters

$$\frac{e}{8m}=0.022\ \text{volt/(gauss-cm)}^2$$

Effect of space charge Space charge of either positive or negative sign can influence the paths of electrons (*see* SPACE CHARGE) Space charge of positive sign is formed by electron beams passing through an imperfectly evacuated space The beam electrons collide with gas atoms and ionize them The heavy ions remain in the path of the electron beam for some time and prevent it from spreading The luminous nodular or thread beams so produced are favorite objects for demonstration

Electron beams in high vacuum on the other hand are subject only to the mutually repulsive forces between the electrons themselves The repulsion is reduced but never canceled by the action of the magnetic fields which surround charges in motion for two electrons moving with the same velocity v parallel to each other the ratio of the magnetic attractive force to the electrostatic repulsive force is given by v^2/c^2 where c is the velocity of light Hence the magnetic force is significant only for electrons of very high energy

The action of the remainder of the electrons in the beam upon any one electron can be approximated adequately by that of a continuous charge distribution equal to the average space-charge distribution The behavior of the edge ray of a uniform circular beam of current I aimed at a point of convergence a distance L from the initial cross section of radius r_B may serve as an example (Fig 3) If the variation of the potential along the axis of the beam is neglected (that is if Φ is assumed to be constant) and the charge density ρ is regarded as uniform within any beam cross section ρ is given by

$$\rho=\frac{I}{\pi r^2}\left(\frac{m}{2e\Phi}\right)^{1/2}$$

and the path equation becomes

$$\frac{d^2r}{dz^2}=\frac{\pi\rho}{\epsilon\Phi}r=\left(\frac{m}{2e}\right)^{1/2}\frac{I}{\Phi^{3/2}}\frac{1}{\epsilon r}$$

Here ϵ is the dielectric constant of vacuum As the result of the repulsive force of space charge the ray under consideration does not cross the axis but reaches a minimum separation r_0 from the axis and diverges from this point on Integration of the differential equation gives

$$r_0=r_B\exp\left[-\epsilon\left(\frac{e}{2m}\right)^{1/2}\frac{r_B^2\Phi^{3/2}}{L^2I}\right]$$
$$=r_B\exp\left[-3.3\times10^{-5}\frac{r_B^2\Phi^{3/2}}{L^2I}\right]$$

For example if $r_B = 1$ mm $\Phi = 10\,000$ volts $L = 10$ cm and $I = 0.001$ amp then $r_0 = 0.037$ mm

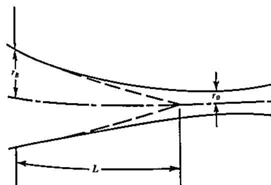


Fig 3 Widening of electron beam by space-charge repulsion as it traverses from left to right

Time varying fields In the preceding discussion it was assumed that the electric and magnetic fields traversed by the electrons were constant in time The total energy of the electron or the sum of the kinetic energy and the potential energy is then a constant Since the potential energy is a function of position only so is the kinetic energy This is no longer true if the fields change appreciably in a period corresponding to the transit time of the electrons For discussion of devices known as betatrons cyclotrons and synchrotrons which utilize this fact to achieve particle energies that are large compared to that imparted by any applied voltage *see* PARTICLE ACCELERATOR For discussion of electron motion in time varying fields such as are encountered in microwave tubes *see* KLYSTRON MAGNETRON TRAVELING-WAVE TUBE

Electrostatic deflection at high frequencies In the cathode ray oscilloscope the beam deflection ceases to be proportional to the potential difference V applied to the deflection plates if V changes appreciably in the course of the passage of the electron beam through the deflection field If $V = V_0\cos 2\pi ft$ an integration of the transverse impulse impressed on the electron passing between two parallel plates of length l and separation d leads to the following for the deflection angle α

$$\tan\alpha=\frac{V_0 l}{2\phi d}\frac{\sin u}{u}\cos(2\pi ft)$$

Here ϕ is the accelerating potential of the beam and t is the time it passes through the center of the deflection field If f (the frequency of the applied voltage) is measured in sec^{-1} l in centimeters and ϕ in volts

$$u=\pi fl/(2e\phi/m)^{1/2}=5.3\times10^{-8}\,fl/\phi^{1/2}$$

The quantity $(\sin u)/u$ represents the ratio of the deflection sensitivity at frequency f to that at low

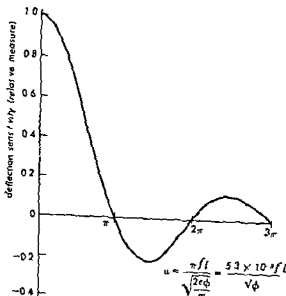


Fig 4 Deflection sensitivity of cathode ray oscilloscope as a function of frequency f

frequencies ($f \to 0$) (Fig 4) Thus for a 10 kilo volt beam and ded respor 82 3% it drop cussio e deflection field is assumed to be a sharply cutoff uniform field with the effects of fringe fields neglected [F C RA]

Bibliography K R Spangenberg *Vacuum Tubes* 1948 V K Zworykin G M Morton E G Ramberg J Hillier and A W Vance *Electron Optics and the Electron Microscope* 1945

Electron optics

The branch of physics concerned with the motion of free electrons under the influence of electric and magnetic fields The term electron optics is derived from the fact that the laws governing electron paths in such fields are formally identical with those governing light rays in media of varying refractive index Both may be derived from Fermat s law This law states that the actual light ray or electron path passing through two prescribed points A and B is that which makes the integral

$$\int_A^B n\,ds$$

carried out over it a minimum The refractive index n is for electrons

$$n = \sqrt{\phi + \frac{2e\phi^2}{mc^2}} - \sqrt{\frac{e}{2m}}\,A \cos \chi$$

Here $-e/m$ is the specific charge of the electron c the velocity of light ϕ the accelerating potential of the electron A the magnetic vector potential and χ the angle formed by the electron path with the direction of the magnetic vector potential Since for electrons of given kinetic energy ϕ and A are unique functions of position the refractive index n is also a function of position and in the presence of a magnetic field of the direction of the electron path A similar dependence of the refractive index on the direction of a light ray is encountered in crystal optics (*see* CRYSTAL OPTICS)

The study of electron paths analogous to the study of light rays is more properly called geometrical electron optics The electron paths may be regarded as normals to electron waves whose amplitude determines the statistical density of electrons just as the amplitude of a light wave determines the density of light quanta or photons The study of the wave motion associated with electrons is called electron wave optics It describes diffraction and interference effects between electron beams which are in every way similar to the diffraction and interference effects observed with light and x rays

Electron optics finds application in the formation of electron beams as in cathode ray tubes and television camera tubes in the deflection of such beams by electric and magnetic fields and in the formation of electron images as in electron microscopes and image tubes *See* CATHODE RAY TUBE ELECTRON DIFFRACTION ELECTRON LENS ELECTRON MOTION IN VACUUM ELECTROSTATIC LENS MAGNETIC LENS MICROSCOPE ELECTRON [E G R]

Bibliography *See* ELECTRON MOTION IN VACUUM

Electron paramagnetic resonance spectroscopy

The study of magnetic resonance spectra of materials which show paramagnetism because of the magnetic moment of unpaired electrons (*see* ELECTRON SPIN MAGNETIC RESONANCE PARAMAGNETISM) Electron paramagnetic resonance (EPR) spectra are usually presented as plots of the absorption or dispersion of the energy of an oscillating magnetic field of fixed radio frequency versus the intensity of an applied static magnetic field

Electron paramagnetic resonance spectroscopy has been used for detection and identification of paramagnetic materials for determinations of electronic structure for studies of interactions between molecules and for measurements of nuclear spins and moments Among the wide variety of paramagnetic substances to which EPR spectroscopy has been applied are free radicals (including free atoms) impurity centers and compounds of the transition elements rare earths and actinides Electron paramagnetic resonance spectra have been obtained from gases liquids and solids Much of the work has been done with the oscillating magnetic field either in the vicinity of 9×10^9 cps (X band) or $2\,4 \times 10^{10}$ cps (K band) Measurements at other frequencies lying between 10^8 and 10^{11} cps have been performed

Spectra characteristic of individual paramagnetic molecules uncomplicated by magnetic interactions with neighboring paramagnetic molecules may be obtained only from dilute solutions in diamagnetic solvents The required degree of dilution

depends on the nature of the magnetic molecules. Concentrations lower than 10^{16} molecules per cm^3 frequently must be used for solutions of organic free radicals whereas concentrations as high as 10^{19} molecules per cm^3 may sometimes be tolerated for solutions of inorganic ions.

In some cases spin lattice relaxation (exchange of magnetic energy with thermal motions of the environment) obscures the spectra (see MAGNETIC RELAXATION). The effects are especially pronounced in inorganic magnetic ions and frequently require the use of low temperatures. The spectra of organic free radicals on the other hand are not severely affected and may usually be obtained at ordinary temperatures.

Solids Maximum information is yielded by EPR spectra of solid solutions in single crystals. The spectrum of a single species may contain scores of lines their positions and intensities varying with orientation of the specimen relative to the static magnetic field. The many lined structure results in part from interactions of the orbital motion of the electrons with the electric fields of the environment and from interactions of the magnetic moments of the electrons with nuclear magnetic moments. The latter effect (hyperfine interaction) is the sole cause of the splittings in the EPR spectra of most organic free radicals. Most magnetic ions exhibit pronounced anisotropy in their EPR spectra (Fig. 1) because of the anisotropic nature of the orbital (wave functions) which give rise to their magnetism. Most organic free radicals and a few ions with highly symmetrical charge distributions or with highly quenched orbital magnetism exhibit little anisotropy in their EPR spectra.

From analysis of hyperfine interactions with nuclei whose spins and magnetic moments are known details of the distribution of electrons about the nuclei may be determined. The average value of the cube of the reciprocal of the distances between electrons and nuclei the orientation of the orbits relative to the crystal axes and the density of unpaired electrons about the various nuclei in a free radical may be evaluated from the hyperfine structure.

Relative values of nuclear moments of isotopes may be found from the relative splittings produced by them in the same chemical environment. If only one isotope is available its nuclear magnetic moment may be obtained from EPR spectra only if suitable properties of the electronic orbits are known. Nuclear spins on the other hand may be found simply by counting hyperfine components.

Liquids Much work has been done particularly with organic free radicals in liquid solutions where less information is obtainable by EPR spectroscopy than in crystals. Only averages of orienta

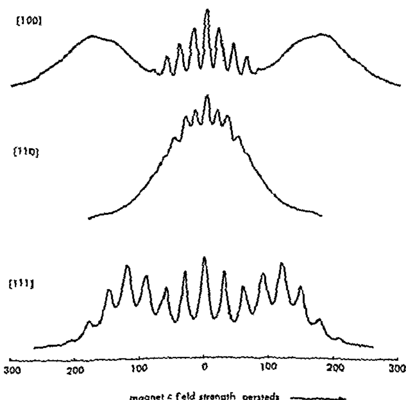


Fig. 1 Spectra showing energy absorption vs magnetic field strength for a dilute solution of FeF_6^{3-} in a single crystal of Na_2KGaF_6. The indices [111], [110] and [100] give the direction of the magnetic field relative to the crystal axes. (Courtesy L. Helmholz)

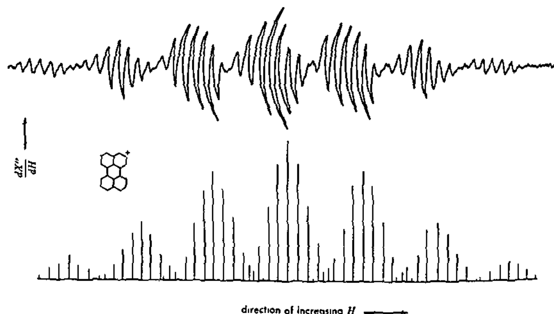


Fig 2 Spectrum of dX''/dH vs H (X'' is the energy absorption, H is the magnetic field strength) for a liquid solution of the free radical perylene positive ion Vertical lines give the positions and intensities of the spectral lines as calculated from molecular orbital theory (Courtesy E de Boer and S I Weissman)

tion dependent properties can be observed in motions in liquids The nature of the average is dependent on the rapidity of the motions Well resolved EPR spectra may be obtained in liquids only if the variations of positions of lines with orientation of molecular axes are not great Many organic free radicals fulfill this requirement and yield lines only a few tenths of an oersted broad in liquid solutions Highly characteristic spectra ranging from those containing only one line (such as the semiquinone of chloranil) to others containing more than 100 lines (triphenylmethyl) have been recorded The spectra of organic free radicals are symmetrical about a center (Fig 2) At fields of 3200 oersteds the centers usually lie within a few oersteds of the position of the resonance of the spin of a free electron that is one not localized at a single nucleus

Hyperfine interactions are responsible for the complexity of the EPR spectra of most free radicals Most of the splittings observed thus far have been produced by H^1, which has a nuclear spin quantum number of ½ ($I = ½$) Splittings by H^2, B^{11}, C^{13}, N^{14} and N^{15} have also been studied The contribution to the splitting by each kind of proton (in a free radical the protons differ in chemical environment) is determined by observation of the EPR spectra of radicals with appropriate substitutions of H^1 by H^2 ($I = 1$)

Rates of electron transfer Migration of electrons among different molecules may produce measurable effects on the EPR spectra In favorable cases electron spin rates and mechanisms may be detected In a stable free radical with resolved hyperfine structure, each line is associated with the frequency of precession of the electron spin in the presence of a particular arrangement of nuclear spins When the electron jumps to a molecule with a different arrangement of nuclear spins its spin precesses at a different frequency Jumps randomly distributed in time with mean frequency $\bar{\nu}$ add breadth $\bar{\nu}$ to the spectral lines as long as $\bar{\nu}$ is small compared with the separation of the various frequencies of precession The method has been applied to measurements of electron transfer reactions with second order rate constants in the range 10^6–10^9 liter/(mole)(sec)

When $\bar{\nu}$ becomes large compared with separation of lines, a new spectrum appears The hyperfine structure of the new spectrum reveals the nature of the groups of atoms which accompany the electron in its migrations See NUCLEAR MOMENTS [S I W]

Bibliography D J E Ingram, *Free Radicals as Studied by Electron Spin Resonance*, 1958, J E Wertz, Nuclear and electronic spin magnetic resonance *Chem Revs*, 55 829–955, 1955

Electron spin

That property of an electron which gives rise to its angular momentum about an axis within the electron Spin is one of the permanent and basic properties of the electron Both the spin and the associated magnetic dipole moment of the electron were postulated by G E Uhlenbeck and S Goudsmit in 1925 as necessary to allow the interpretation of many observed effects among them the so-called anomalous Zeeman effect, the existence of doublets (pairs of closely spaced lines) in the spectra of the alkali atoms, and certain features of x ray spectra See SPIN (QUANTUM MECHANICS)

The evidence for the existence of electron spin is very great and all theory that concerns itself with electronic, nuclear, atomic, and molecular

phenomena includes the electron spin in its formulation to obtain a theoretical structure consistent with experimental observation The addition of spin to the properties of charge and mass that the electron also possesses attributes a rotational motion to the individual electrons See ELECTRON

The spin quantum number is s where s is always ½ (see QUANTUM NUMBERS) This means that the component of spin angular momentum along a preferred direction such as the direction of a magnetic field is $\pm\frac{1}{2}\hbar$ where $\hbar = h/2\pi$ and h is Planck s constant The total angular momentum of the spinning electron is $\sqrt{s(s+1)}\hbar = \sqrt{3}\hbar/2$ however the total angular momentum is not an observable quantity The spin angular momentum of the electron is not to be confused with the orbital angular momentum of the electron associated with its motion about the nucleus In the latter case the maximum component of angular momentum along a preferred direction is $l\hbar$ where l is the angular momentum quantum number and may be any positive integer or zero The total orbital angular momentum is $\sqrt{l(l+1)}\hbar$ In the following discussion the angular momentum or the magnetic dipole moment will be terms used to describe the maximum component of these quantities along a field direction

Electron magnetic moment The electron has a magnetic dipole moment by virtue of its spin The approximate value of the dipole moment is the Bohr magneton μ_0 which is equal to $eh/4\pi mc = 9.27 \times 10^{-21}$ erg/oersted where e is the electron charge measured in electrostatic units m is the mass of the electron and c the velocity of light (see MAGNETON) The orbital motion of the electron also gives rise to a magnetic dipole moment μ that is equal to μ_0 when $l = 1$ (Fig 1) The direction of the dipole moment is in each case opposite to that of the angular momentum (both are vectors) and the magnetic moments are therefore negative For a positron (a positively charged particle having the same mass and magnitude of charge as the negatively charged electron) the magnetic moments are positive that is in the same direction as the angular momentum See POSITRON

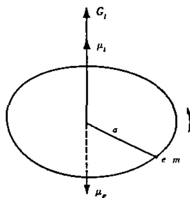


Fig 1 Diagram illustrating the orbital angular momentum G_l and orbital magnetic moment μ_l due to a negative charge revolving in a circle of radius a

The orbital magnetic moment of an electron can readily be deduced with the use of the classical statements of electromagnetic theory in quantum mechanical theory the simple classical analog of a current flowing in a loop of wire describes the magnetic effects of an electron moving in an orbit The spin of an electron and the magnetic properties associated with it are however not possible to understand from a classical point of view The classical radius of the electron is $e^2/(2mc^2) = 1.41 \times 10^{-13}$ cm a distribution of mass and electric charge within this radius leads to the calculation of a peripheral velocity of the electron far greater than the velocity of light which is of course wholly precluded by the statements of the special theory of relativity No theory of the structure of the electron has been formulated which makes the spin of the electron amenable to simple pictorial understanding Nevertheless the interpretation of the spectra of atoms and molecules the magnetic properties of materials and other phenomena on an atomic scale unambiguously require that the electron *have* the specified properties To the extent that physical theory assumes these properties and does not concern itself with questions about the structure of the electron it is quite adequate to deal with a large range of physical phenomena in a highly quantitative way

The Lande g factor g is defined as the negative ratio of the magnetic moment in units of μ_0 to the angular momentum in units of $\hbar$ For the orbital motion of an electron $g_l = 1$ For the spin of the electron the appropriate g value is $g_s \cong 2$ that is unit spin angular momentum produces twice the magnetic moment that unit orbital angular momentum produces

The following discussion is limited to atoms which have a single electron outside of closed electron shells Both the orbital and spin angular momenta of the electrons within closed shells add up in such a way that their net angular momentum is zero The single electron outside closed shells may have $l = 0\ 1\ 2$ By the usual rules developed from quantum mechanics the total angular momentum quantum number of the electron which is called j is $l \pm s$ (Fig 1) except when $l = 0$ in which case the total angular momentum is s For instance when $l = 1$ $j = ½$ or $\frac{3}{2}$ These relations may be represented by vector diagrams as shown in Fig 2 Since the rotation of the electron about the nucleus produces a magnetic field at the electron

levels in all single electron atoms except when $l = 0$ in which case level is single In the case of sodium the familiar yellow lines (the D lines) comprise a closely spaced doublet that arises from a transition from a state of $l = 1$ to one of $l = 0$

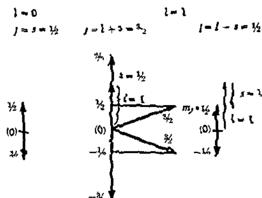


Fig 2 Diagrams illustrating addition of l and s vectors into a resultant j vector The allowed values of m are also shown In no case does the bracketed (0) give an allowed value

The doubling of the lines is a direct consequence of the existence of the electron spin

Energy level splitting The total electronic magnetic moment of an atom depends on the state of coupling between the orbital and spin angular momenta of the electron (see Fig 2) In the single electron case an atom in the state for which $l = 0$ has only the magnetic moment associated with the spin and this moment may be oriented in either of two directions with respect to an externally applied magnetic field However when the atom is in a state for which $l = 1$ j can be ½ or 3/2 and the Lande g factors for these two states are ⅔ and 4/3 respectively That is the magnetic moment per unit angular momentum is equal neither to that which characterizes the orbital motion nor to that which characterizes the spin motion In a magnetic field a single energy level characterized by l and j is split into several energy levels each described by the component of j m_j along a magnetic field where $m_j = j$ $j - 1$ ... $-(j-1)$ $-j$ The energy of the level is the zero field energy pl s the ...

... between the two levels plus the difference between two magnetic energy terms as given previously The resultant splitting of the line which is single at zero magnetic field into a line of two or more components is called the Zeeman effect The Zeeman effect has been called anomalous when it is not explicable purely in terms of the orbital motion of the electron The introduction of the electron spin allows the interpretation of all observed Zeeman effects See ZEEMAN EFFECT

Atomic beam measurements Prior ... to ...

... u is derable natural width Exact measurements of the splitting of lines in a magnetic field therefore could not be made on such lines and all data on the Zeeman effect were consistent with the statement that $g = 2g$ With the development of spectroscopy by the atomic beam method a new order of precision in the measurement of the frequencies of spectral lines became possible Consider again a single electron atom in a state characterized by some value of l It is possible to observe transitions between levels characterized by single values of l s and j but by different values of m_j In general m must change by ±1 in such a transition The frequency of such a spectral line is then f $g_j\mu_0 H/h$ and for readily available laboratory fields up to about 10 000 oersteds the frequency lies in the range that can be generated by electronic means In principle if f is found at a known H g can be determined However a measurement of H to high precision is extremely difficult especially under the experimental requirement that f must be measured simultaneously Suppose however that one has available two different atoms (say 1 and 2) in states of different j and measures the frequency of transition in the same magnetic field Then $f_1/f_2 = g_{j1}/g_{j2}$

Now g_j is a linear combination of g and g_s $g_j = \alpha g_l + \alpha_s g_s$ The constants α and α_s can be calculated with high accuracy for many atomic systems The experimental ratio f_1/f_2 then yields directly the ratio g_s/g_l This quantity has been found on the basis of the experiments described here in principle (in practice they are of considerably greater complexity) and on the basis of other experiments to be $g_s/g = 2(1\,001168 \pm 0\,000005)$ The magnetic moment of the electron is therefore not μ_0 but $1\,001168\ \mu_0$ It is only through the refined experimental techniques of spectroscopy by the method of atomic beams that it has been possible to obtain the precise data that lead to an accurate knowledge of the magnetic moment of the electron For a detailed discussion of these techniques see MOLECULAR BEAMS

It is not possible to give a qualitative description of the effects which give rise to the deviation of the spin magnetic moment of the electron from μ_0 The detailed theoretical calculation of the quantity is in the domain of quantum electrodynamics and involves the interaction of the zero point oscillations of the electromagnetic field with the electron The calculated value of the spin moment of the electron is in excellent agreement with the experiment ... see ...

Bibliography F K Richtmyer E H Kennard and T Lauritsen *Introduction to Modern Physics* 5th ed 1955 V F Weisskopf Recent developments in the theory of the electron *Rev Mod Phys* 21(2) 305–315 1949

Electron tube

A generic term given to a large family of devices Electron tubes include all partially evacuated tubes whose characteristics are derived from the flow of electrons through the tube In the general category of electron tubes are found the two subclasses vacuum tubes and gas tubes In England and other

Commonwealth countries the term valve is used in tead of vacuum tube

Vacuum tubes are tubes which are almost completely evacuated of gas so that electron flow occurs in what is essentially a vacuum Gas tubes conta n a chosen gas at a low pressure so that the electron flow is accompanied by gas ionization

Electron tubes are of tremendous importance in our modern technology In 1958 approximately 600 000 000 vacuum tubes of one kind or another were manufactured and used in the United States They are the basis for modern communications rad o and television They are extensively used in computers and industrial applications The defense activities of the United States consume probably at least one-third of the total output The reason for this widespread use is that the electron tube is a versatile device which readily performs a number of functions notably those of amplification oscillation switching detection and frequency conver s on which are not readily performed by other devices

Vacuum tubes exist in numerous forms These are primarily low power devices but they also exist in h gh power form for transmitters and other purposes They may be either simple diodes that is two-electrode tubes or they may be multielectrode tubes capable of performing a wide range of functions Vacuum tubes are primarily used in applications where low noise and high frequency are involved. In contrast gas tubes are used for high current low frequency applications They may be e ther simple diodes which are used primarily as rectifiers or control type tubes having three or more electrodes for a variety of purposes The gas tubes are used mainly in industrial applications where high power handling ability overshadows their frequency limitations

Some of the phenomena associated with present day electron tubes were first noticed in the latter part of the last century In 1883 T A Edison observed some peculiar effects in light bulbs which were later recognized as due to electrons H Hertz observed photoelectric emission in 1887 W C Roentgen observed x rays in 1895 The electron itself was probably first identified by J J Thomson who in 1897 also measured its ratio of charge to mass Probably the first electron tube was a cathode ray tube built by K F Braun in 1897 A Wehnelt discovered oxide emission in 1903 and this led to the development by J A Fleming of the vacuum diode in 1904 With the invention of the vacuum triode by L deForest in 1906 the age of electronics was ushered in The tetrode was developed in 1919 by W Schottky and the pentode by G Jobst and B D H Tellegen in 1926 A W Hull developed the thyratron in 1929 and the magnetron in 1921 R Varian and S Varian invented the klystron in 1938 and R Kompfner introduced the travel ng wave tube in 1946

For further information *see* CATHODE RAY TUBE GAS TUBE INFRARED IMAGE CONVERTER TUBE MICROSCOPE ELECTRON MICROWAVE TUBE NUMBER INDICATOR TUBE PHOTOTUBE SWITCHING TUBES VACUUM TUBE X RAY TUBE [K R S]

Electron volt

A unit of energy used for convenience in atomic systems Specifically it is the change in energy of an electron or of any particle having a charge numerically equal to that of an electron when it is moved through a difference of potential of 1 mks volt Its value (in mks units) is obtained from the equation

$$W = qV$$

where W is energy in joules q the charge in coulombs and V the potential difference in volts For a potential difference of 1 volt and the electronic charge of $1\,601 \times 10^{19}$ coulomb the electron volt is $1\,601 \times 10^{19}$ joule *See* ELECTRON IONIZATION POTENTIAL [G H MI]

Electronegativity

Defined by Linus Pauling as the power of an atom in a molecule to attract electrons to itself The concept is used by chemists to designate the relative electropositive or electronegative character of an element as it appears in a given state of chemical combination Most of the elements are capable of exhibiting different valences or states of oxidation and they also may form chemical bonds of different types Thus electronegativity cannot of itself be considered a precise or invariant property of an element but rather is a more general characteristic which becomes very useful in discussions of reaction mechanisms and bond polarities

At first view it would seem that electronegativity has somewhat the same meaning as electron affinity or the attraction of an atom for an additional electron However electron affinities are most readily calculated from the Born Haber thermochemical cycle which relates lattice energy to the ionization

heat of dissociation of the nonmetal from its molecular form and the heat of formation of the compound from atoms of metal and nonmetal It follows that the electron affinity represents energy that results from the addition of a single electron to an isolated atom of an element capable of attracting that electron a situation quite far removed from the usual problems of polarity of covalent bonds and their consequent modes of reaction For this reason the concept of electronegativity inexact and unspecific as it may seem is still a very

or upon thermochemical data The original scale proposed by Pauling in 1932 was based upon the difference between the bond energy of a compound AB and the mean of the values for the completely homopolar bonds A—A and B—B The A—B

energy exceeds the geometric mean of the A—A and B—B bonds to an increasing extent as the elements A and B diverge in relative positive or negative character The square root of the difference in bond energy was shown by Pauling to provide a convenient scale of relative electronegativity of the bond partners In more exact terms if Δ represents the difference in bond energy of the A—B bond over the average of the A—A and B—B bonds then 0 208 times the square root of Δ represents the difference in electronegativity between the elements A and B (the factor 0 208 is used to convert energies in kilocalories to electron volts) The scale then is determined principally by the values for carbon (2 5) nitrogen (3 0) oxygen (3 5) and fluorine (4 0) from combination of these elements with others and a consideration of the values derived from the expression 0 208 times the square root of Δ the electronegativities for many other elements were determined Table 1 gives some of the values as they appeared in the Pauling scale

R Mulliken reasoned in 1935 that if the first ionization energy of an atom (that is the energy required to remove one electron from the atom to an infinite distance) were known and if the electron affinity could be calculated from information available about crystalline compounds of the element then the average of the ionization energy and the electron affinity was proportional to the electronegativity according to the Pauling scale In order to bring into coincidence the scales derived by the two methods of calculation the sum of ionization energy and electron affinity is divided by 130 giving the values 4 0 for fluorine 3 0 for chlorine 2 1 for hydrogen 1 0 for lithium and 0 9 for sodium In the last two instances the electron affinity of the alkali metal was assumed to be zero in accordance with the chemical properties of these elements

Other methods for calculating relative electronegativities have been suggested by W Gordy M L Huggins R T Sanderson and H O Pritchard references to some of their works will be found in the bibliography Their methods bring in such diverse quantities as bond stretching force constants dipole moments covalent radii and heats of formation as bases for the suggested values A L Allred has gone further with a proposal for relating electronegativity to the electrostatic force of a nucleus upon the surrounding electrons in a compound as given by the ratio of the effective nuclear charge to the square of the atomic radius When adjusted in scale, all of these methods converge to a considerable degree in that they provide relative electronegativities not much different from those given above in the table of Pauling values Perhaps the only significant exception is that later methods of calculation indicate an alternation of electronegativity for elements in the IVb and Vb subgroups Thus although the Pauling values for electronegativities of the IVb elements were carbon 2 5, silicon 1 8, germanium 1 7 and tin 1 7 Sanderson later arrived at the values 2 47 1 74 2 31 and 2 03 respectively By a variety of physical measurements Allred then proposed the values 2 60 1 90 2 00, and 1 93 respectively This alternation is clearly shown in many chemical reactions of germanium which are not shown by silicon or tin As the methods for calculation assume a broader basis it may be expected that other variations will become apparent, and that these in turn will be related to observed chemical properties.

The revised values of electronegativity for some selected elements are given in Table 2 Although these apply strictly only to specific orbital configurations and so are not wholly consistent with the scale represented in Table 1 they may be more helpful in particular situations See IONIC CRYSTALS, MOLECULAR STRUCTURE AND SPECTRA [E G R]

Table 1 Pauling's values of electronegativity 1940

Cs	0 7	Be	1 5	Se	2 4
Rb	0 8	Ti	1 6	I	2 4
K	0 8	Zr	1 6	C	2 5
Ba	0 9	Ge	1 7	S	2 5
Na	0 9	Sn	1 7	Br	2 8
Sr	1 0	Si	1 8	Cl	3 0
Ca	1 0	Sb	1 8	N	3 0
Li	1 0	B	2 0	O	3 5
Mg	1 2	As	2 0	F	4 0
Sc	1 3	Te	2 1		
Al	1 5	H	2 1		

Table 2. Revised values of electronegativity 1958

H	2 20	Ge	2 00
C	2 60	As	2 10
N	3 05	Se	2 55
F	3 90	Br	2 95
Si	1 90	Sn	1 93
P	2 15	Sb	2 05
S	2 60	Te	2 30
Cl	3 15	I	2 65

Bibliography A L Allred *J Inorg & Nuclear Chem* 5(4) 264–288 1958, S Glasstone *Textbook of Physical Chemistry*, 2d ed 1946 W Gordy *Phys Rev*, 69 604, 1946, M L Huggins, *J Am Chem Soc* 75 4123 1953, L Pauling *Nature of the Chemical Bond* 2d ed 1940, R T Sanderson, *J Chem Educ* 29 539–544 1952 and 31 2 7 1954

Electronic countermeasures

A term used to describe a variety of military equipments and techniques which are used to detect, disrupt or deceive the radio communications radar and guided missiles of an enemy

Electronic countermeasures (ECM) have assumed crucial significance in modern warfare because of the widespread use of radio for speedy communications radar for early warning and the growing use of guided missiles many of them guided to their targets by radio or radar

The wide variety of ECM equipment and techniques can be classified into three general categories reconnaissance active and passive

Reconnaissance ECM equipment is used to detect and analyze electromagnetic radiation from

radio and radar transmitters in enemy aircraft missiles ships and fixed installations Passive ECM devices and techniques change the nature of the energy reflected back to enemy radars they generate no electromagnetic energy themselves Active ECM devices generate electromagnetic radiation designed to interfere with enemy radio or radar signals or to confuse enemy radio or radar operators Active ECM sometimes is called jamming For a discussion of radar and radio techniques and equipment *see* RADAR RADIO

Reconnaissance Reconnaissance ECM systems usually carried in ferret aircraft submarines or ships, consist of one or more extremely sensitive radio or radar receivers which can be rapidly tuned over a wide portion of the electromagnetic spectrum in search of enemy transmissions

When electromagnetic radiation is intercepted automatic direction finding techniques are employed to pinpoint the bearing to the transmitter If the signal comes from an enemy radar it is recorded on magnetic tape Subsequently the tape is played back through suitable equipment which analyzes the radar signal to determine its pulse repetition rate pulse width and other important characteristics

Electronic reconnaissance is also valuable in spotting the coverage of an enemy radar and in locating any blind spots in its coverage due to terrain obstructions

Another type of ECM reconnaissance receiver is carried aboard modern bombers to warn the pilot when his airplane is being illuminated by radar energy from the ground or from an interceptor missile or airplane This enables the pilot to take evasive maneuvers or to turn on the bomber s own active ECM

Generally speaking a vehicle equipped with an ECM reconnaissance receiver can detect radiation from an enemy radar at considerably greater distance than the radar can detect the presence of the vehicle The reason is that the strength of the energy striking the vehicle is always considerably greater than that which is reflected back to the radar

Passive ECM The first passive type of ECM to be used against radar and the best known is called chaff It consisted of thousands of tiny strips of tinfoil which World War II bombers dumped overboard during flights over Germany (Fig 1) Electromagnetic energy from the German antiaircraft radars was reflected back from the thousands of

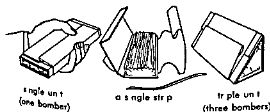


Fig 1 Chaff (Aviation Week)

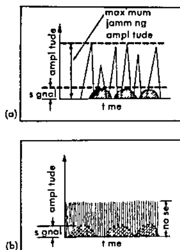


Fig 2 Jamming techniques (*a*) Spark jamming (*b*) White noise (*Aviation Week*)

strips of tinfoil producing spurious echoes on the radar scopes obscuring the echoes from Allied bombers and badly confusing the radar operators

Because today s bombers fly at vastly higher speeds it is easier for radar operators to tell the difference between slow moving echoes produced [illegible]

Another type of passive ECM is the corner reflector or Luneberg lens These are simple mechanical devices which tend to reinforce or strengthen the radar energy which they reflect back to the radar If one or more of these devices are installed for example on a small airplane the energy reflected back to the radar is so enhanced that the airplane appears to be a large bomber on the radar scope Tiny drone aircraft equipped with corner reflectors can be launched from a bomber to produce multiple targets on an enemy radar scope so that the radar operator does not know which is the real bomber or to confuse an attacking missile which uses radar to find its target

Active ECM The earliest form of active ECM was employed against radio communications in an effort to blot out or overpower an enemy transmission or to so irritate the radio operator that he could not do an efficient job This brute force approach sometimes called jamming requires only comparatively simple equipment but has the disadvantage that it instantly alerts the enemy to the fact that he is being jammed

The following are representative of the type of [illegible]

receivers by some electric razors Spark jamming

is relatively easy to produce with simple equipment at low frequencies in the radio spectrum

White noise This is a more sophisticated version of spark jamming and can be used at any frequency In this technique the random noise produced by a gas discharge tube is used to modulate the output of the ECM transmitter White noise and spark jamming have the disadvantage of requiring large powerful transmitters because energy is radiated over a fairly large portion of the spectrum *See* NOISE ELECTRICAL

Sweep through A moderately simple but effective type of jamming can be achieved by sweeping the ECM transmitter s carrier frequency back and forth over a portion of the radio spectrum several hundred times per second This creates hundreds of noise pulses per second in the enemy s radio receiver each momentarily blanking out the incoming message Because of the time required by the radio receiver detector circuits and the human ear to recover from each of these pulses the effect is practically equivalent to continuous jamming Less transmitter power is required because the ECM energy is concentrated in a relatively narrow part of the radio spectrum at any instant Another advantage is that a single jammer of this type can simultaneously be employed against a number of enemy receivers each operating at a different frequency

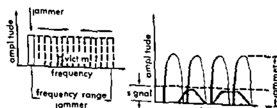


Fig 3 Sweep through jamming (*Aviation Week*)

Heterodyne If an ECM transmitter operates on a carrier frequency that differs only slightly from that of the enemy transmitter receiver it can produce a continuous beat note in the receiver output This high pitched squeal similar to that heard in home broadcast receivers that are in need of adjustment can be quite irritating to an enemy operator If the ECM transmitter carrier frequency is slowly varied back and forth across the enemy radio carrier frequency the resulting beat note in the radio operator s ear sounds like the proverbial wail of a banshee

Radar jamming techniques Techniques used for jamming radio transmissions were quickly adapted for use against radar when radar first appeared in World War II Because radar receivers are designed to operate from extremely weak signals reflected from the target only low power ECM transmitters are required to swamp the radar echo signal completely

If the ECM transmitter is modulated by white noise it tends to obscure the echo from the target much as tall grass hides a golf ball To counter the effectiveness of ECM, radar designers have developed techniques intended to discriminate between repetitious signals such as are received from a target and random ones produced by such simple ECM transmitters This, plus the fact that it is better to confuse or deceive the radar operator without letting him know that he is being jammed has led to the development of far more sophisticated types of active ECM

One approach is to create a number of false targets on the radar scope Radar has two basic means for establishing target bearing and distance Bearing is determined by means of a device that indicates the direction the radar antenna is pointing when the echo is received from the target Distance to the target is determined by measuring the time it takes for a pulse of radar energy to travel from the antenna to the target and back to the antenna

If the target say a bomber, carries a small ECM transmitter that sends out a series of pulses at suitably spaced intervals each time that a pulse from the enemy radar is received the enemy radar will indicate the presence of a group of targets each at a slightly different range and will be unable to determine which is the real target The pulses transmitted by the ECM equipment naturally must be identical to those transmitted by the enemy radar in terms of their shape time duration repetition rate and radio frequency Radar directed interceptors and guided missiles require accurate information on the distance to the target in order to compute the flight path for interception of the target

If the target s ECM transmitter sends out a single pulse every time it receives a burst of radar energy from the interceptor or missile then slowly begins to shift the timing of its own pulse transmissions the tracking circuits of the interceptor or missile radar will measure a target range that is different from what it actually is This can cause the interceptor or missile to compute and fly an erroneous path thus missing the target

Decoys A more effective way to divert an attacking interceptor or missile is to mislead it as to the bearing of the target This can be accomplished by means of decoys such as small drone aircraft outfitted with passive reflectors as previously described or with more sophisticated ECM active transmitters

A tiny drone missile outfitted with an ECM transmitter makes a decoy which is a much more attractive target to an enemy radar than the real target because the signal transmitted by the ECM equipment is much more powerful than the echo reflected from the real target If the bomber releases an ECM decoy missile from its wingtips or bomb bay and guides it by remote control on a different course from the bomber s the enemy radar will lock onto and follow the decoy instead of its desired bomber target

Counter ECM Military emphasis on electronic countermeasures has spawned the field of counter

ECM or techniques for foiling enemy ECM equipment

Many of the passive and active ECM devices must be tuned accurately to the operating frequency of the radio or radar they are designed to counter One way to negate their effectiveness is to design radio and radar systems so that their operating frequencies can be changed almost instantly To prevent radars from being fooled by ECM mimics new radars are designed to permit them to change the characteristics of their pulses as well as their operating frequencies rapidly

Although it is common practice to speak of jamming a radar as if electronic warfare were merely a battle between two inanimate objects in reality electronic warfare is a continuous battle of wits between people armed with ECM and counter ECM equipment

The scientists and engineers who develop such equipment are also engaged in a continuous battle When the ECM engineers come up with a new technique which can jam or fool the best existing radar the radar designers bend their efforts to make their sets immune to the new technique The result is a never ending race between designers of ECM and counter ECM equipment and techniques

In modern warfare electronic countermeasures the silent, highly secret weapons of electromagnetic detection deception and disruption have become a vital factor [P J K]

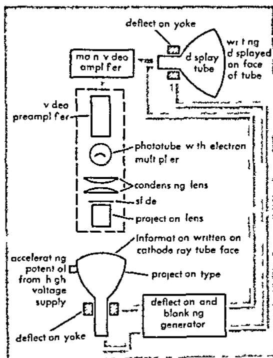
Electronic writing

The use of electronic circuits and electron devices to reproduce symbols such as an alphabet in a prescribed order on some electronic display device for the purpose of transferring information from some source to the viewer or user of the display device

The flying spot scanner. The most general although cumbersome electronic writing system makes use of a flying spot scanner suitable scanning and video amplifier circuits and a display tube which may be a cathode ray tube with a long persistence screen or a storage tube as shown in the illustration

The raster produced by the scanning beam at the scanner tube surface may be similar to that of the television system covering essentially the entire screen The information to be transferred is written or printed on the face of the tube and therefore modulates the light reaching the phototube pickup Alternately information may be preprinted on a transparent slide and inserted in the optical system as shown The video signal contains the information to be transmitted The spot size required number of scanning lines and video bandwidth all affect the accuracy of reproduction *See* CATHODE RAY TUBE TELEVISION

Alternatively a preprinted alphabet may be placed on the scanner tube screen The scanning spot may then be switched to the desired character according to some specific program to scan a single character at a time The deflection waveforms



The flying-spot scanner for electronic writ ng

are then less critical but the total writing speed is slower The scanning sequence of position at the reproducer tube is in accordance with the specified program

Cathode ray tube reproduction Essentially all of the functions of writing described for the flying spot scanner and its associated circuits and display device are combined in the character reproducing cathode ray tube (*see* STORAGE TUBE) In principle a character matrix with holes in the shape of the characters of the alphabet to be used is in front of the electron gun in a cathode ray tube The spot is large enough to cover one complete character and is deflected from one character to another by voltages or currents applied to the deflection system in accordance with a prescribed program The beam passing through the character aperture assumes the shape of the character and is then deflected to a programmed point on a phosphor or storage screen With the character writing cathode ray tube no scanning other than that required to position the beam on the desired character is required

Character producing waveforms Any character may be written on the screen of a cathode ray tube by the application of the appropriate combinations of waveforms applied to the horizontal and vertical deflection systems in the manner of the Lissajous figure A point by point plot of a series of x y coordinates corresponding to points on the outline of the characters can be made and the sequence varied until the simplest possible time sequence is found Waveforms of voltage corresponding to these plots for specific symbols can then be synthesized If these waveforms can be produced elec

tronically in accordance with a specific program and applied to the *x* and *y* deflection systems the characters may then be written on the face of the tube and positioned in the desired position

For a complete writing system an electronic programmer together with various control circuits for generating the required waveforms in proper time sequence must be included *See* WAVEFORM NONSINUSOIDAL [G M C]

Electronics

The branch of science and technology relating to the conduction of electricity through gases or in vacuum Electronics is generally considered to be the study and application of electron motion including the means for producing it the laws governing it and the means for controlling it for useful purposes By this definition electronics embraces a broad field of intellectual and industrial effort without clear boundaries

More narrowly electronics is concerned with the design manufacture and application of electron tubes Electron tubes are found in such diverse applications as home entertainment by radio and television communication by wire or wireless detection location and control of aircraft distribution and control of electric power production of x rays control of industrial processes and all aspects of national defense The newer field of solid state electronics has produced transistors semiconductor diodes and other solid state devices which are used in many of the same applications as electron tubes *See* DIODE SEMICONDUCTOR TRANSISTOR It is difficult to find aspects of human endeavor that are not touched in some way by electronics

History All electronic devices have resulted from a by product of Thomas Edison s research on the incandescent lamp He discovered in 1883 that a weak electric current would flow across a partial vacuum between a heated filament and a cold metallic electrode This current would flow in only one direction if the potential between the filament and the plate were reversed in polarity the current would cease It was clear that the carriers of the electricity were electrically charged Actually they were electrons

The first practical use of the Edison effect was made by J A Fleming in 1897 He utilized the unidirectional property of the electron carried current to form a detector of wireless signals The Fleming valve was the prototype of the modern diode tube now widely used as a power rectifier or as a radio detector

A tremendous step forward in the utilization of electron motion was made in the triode invented by Lee De Forest in 1907 He introduced a third electrode (grid) into the two element Fleming valve This electrode situated between the heated cathode (the source of electrons) and the anode the receptor of the electrons has the important ability to control the flow of electrons through the tube by the application of small voltage changes Because the output (controlled) voltage is vastly greate than the voltage required by the grid to control t the triode is an amplifier It is most widely used as a voltage amplifier, but may also serve as a current amplifier and power amplifier

Modern pentodes and other multielement tube are modifications of De Forest s triode for special purposes or characteristics See ELECTRON TUBE

Growth of electronic industry For many years after the invention of the triode electron tube were not widely used During World War I there was some application in radio communication bu the present $1 000 000 000 electronics industry actually began with the advent of broadcasting in 1922 That year 1 000 000 electron tubes were sold at an average price of $6 and 100 000 receivers were sold at an average price of $50 By 1925 these figures had jumped to 12 000 000 tubes at $2 4 and 2 000 000 receivers at $82 in 1929 the figures were 69 000 000 tubes at $2.50 and 4 200 000 receivers at $110 From 1922 to 1929 total annual sale of electronic equipment rose from $60 000 000 to $842 000 000 During the early years of this period many of the sales were to individuals who made their own radio receivers After this time the bul of sales of parts and accessories was to manufacturers

The electronics industry now produces abou 450 000 000 tubes 13 000 000 radio receivers and 6 000 000 television receivers per year Approximately 350 000 production workers are employed by the electronics industry

The importance of electronics to military and civilian aircraft has increased steadily from World War I when the total capital investment for electronics in all civil aircraft was about $400 Jus prior to World War II the average fighter plan carried about $3000 in electronic equipment h average bomber about $5000 Today the capital investment in electronic equipment on a *DC 6* commercial airliner is roughly $30 000 On high performance military fighters the figure is *close* t $300 000 per aircraft and jet bombers *carry* nearl $6 750 000 in electronic apparatus In 1953 *aiming* equipment for antiaircraft guns required *500* tube and 20 000 electronic parts *See* NAVIGATION SYSTEMS ELECTRONIC

Science of electronics Many materials will emi electrons which are then free to move through a vacuum or a gas There are three types of electron emission (1) thermionic in which materials give off electrons when heated (2) photoelectric in which materials release electrons *when irradiated* by light (3) field emission when *electrons* are freed by high potential fields *See* ELECTRON EMISSION

In a vacuum electrons travel in straight lines in a gaseous tube the vacuum is only *partial* and electrons collide with atoms of the gas ionizing the atoms and knocking other electrons *from them See* ELECTRICAL CONDUCTION IN GASES ELECTRON MOTION IN VACUUM

Electrons are also free to move and conduct electricity in certain semiconductors like selenium silicon or germanium See SEMICONDUCTOR

Technology of electronics The many jobs performed by electron tubes depend upon three basic functions rectification (in diodes) and amplification and oscillation (in multielement tubes) The various ways in which these functions are performed by the tubes and the accessory circuitry constitute the electronic art and science See CIRCUIT ELECTRONIC

Communications applications In a typical radio receiver amplifier tubes receive incoming energy at a level of a few microvolts and increase it to a level of about one volt At this point the desired information is extracted from the amplified current by means of a tube acting as a rectifier Further amplification produces the power needed to actuate the loudspeaker In the superheterodyne type of receiver an oscillator is employed to change the frequency of the incoming wave to a more easily amplified lower frequency In addition rectifier tubes change the alternating current available in homes to high voltage direct current required by the tubes At the transmitting station the same tube functions are used to combine the output of a microphone with a high frequency wave which acts as a carrier See COMMUNICATIONS ELECTRICAL

Industrial applications The extreme versatility speed and accuracy of electronic circuits have resulted in their application to industrial and control circuits With electronic components in the control circuits only small currents are used and little power is dissipated in the control circuits Electronic control has enabled manufacturers to make products better faster and at less cost See CONTROL SYSTEMS

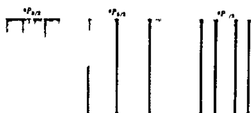
The extreme sensitivity of the electron tube is often used to perform jobs that could not be done in any other way Tubes are available which will measure electric currents as low as 10^{-14} ampere and use them to control a process involving kilowatts of power The same tubes are used in medical research and in laboratories where measurements are needed and where the process being measured must not be influenced by the measuring tool See ELECTRICAL MEASUREMENTS INSTRUMENTATION

[K HE]

Electrooptics

That branch of physics which deals with the influence of an electric field on optical phenomena This influence is exerted through matter which emits absorbs or transmits light The chief electrooptical effects are electrooptical birefringence or the Kerr effect and the Stark effect There are numerous other effects which can be connected with the linear or quadratic Stark effect

An applied electric field may induce an electric dipole and thus change optical transition probabilities In particular otherwise forbidden lines may appear in absorption or emission spectra



Splitting of absorption lines by the electric field in the crystal gadolinium acetate ($GdAc_3 \cdot 6H_2O$) (Johns Hopkins University)

In many cases the electric field is an internal one caused by the presence of neighboring ions in a gas liquid or solid This kind of field except in crystals at low temperatures varies in time and space It causes broadening and shifts of spectrum lines From a study of such effects the ion density can often be determined

In some crystals which have sharp absorption lines at low temperatures the electric field in the vicinity of an absorbing ion is not constant in space This causes a shift and splitting of the absorption lines which may be quite large as shown in the illustration A study of such effects reveals important information on the nature of the electric forces in crystals See KERR EFFECT STARK EFFECT

[G H D]

Electroosmosis

The movement in an electric field of liquid with respect to colloidal particles immobilized in a porous diaphragm or a single capillary tube The phenomenon of electroosmosis is the converse of streaming potential See STREAMING POTENTIAL

Electroosmosis was first observed by F Reuss in 1809 This investigator pushed two vertical glass cylinders into a mass of wet clay filled the cylinders with water and inserted metal electrodes When an electrical potential was applied the level of water rose at the negative electrode but fell at the positive pole to which suspended clay particles were attracted These results were in accord with the view that the clay particles were negative with respect to the water In 1816 R Porret made similar observations employing membranes of animal origin G Quincke in 1862 carried out electroosmosis in single capillary tubes A porous diaphragm can be considered as a multiplicity of capillary tubes

Wiedmann s first law (1852) deduced from experimental studies may be stated as follows $V = \zeta ID/4\pi\eta\kappa$ where V is the volume of liquid transported per second I is the current D is the dielectric constant η is the viscosity and κ is the specific conductance and ζ is the zeta potential

It is possible to combine this expression with Poiseuille s law ($V = \pi P r^4/8\eta l$ where P is the pressure forcing the liquid through the capillary l is the distance between the electrodes and r is the capillary radius) to give $P = 2\zeta ED/\pi r^2$ where

E is the applied potential Thus the difference in hydrostatic pressure maintained across a porous membrane or capillary is (1) directly proportional to the applied potential (2) inversely proportional to the square of the capillary radius and (3) independent of the cross section of the diaphragm

Electroosmosis has been applied to the removal of water from peat and moist clays and in the drying of dye pastes Attempts to use electroosmosis to aid in the tanning of leather have not met with success Indeed the technical success of electroosmosis has been very limited doubtless as a consequence of its slowness and the results do not justify the cost of the electrical power required See ELECTROOSMOTIC DEWATERING ELECTROPHORESIS OSMOSIS [W O M]

Electroosmotic dewatering

A method of drying out an excavated area used to increase the strength of the ground and permit tolerable excavation slopes This is achieved by forcing the flow of ground water to drainage wells by the application of a direct current between electrodes inserted in the ground Electroosmotic dewatering is confined to fine grained soils such as silts and clays where frictional forces prevent the natural flow of ground water to drainage wells For a discussion of dewatering in coarse grained soils see WELLPOINT SYSTEMS

The ground water between the fine grains is made up of three layers an uncharged inner core a negatively charged layer fixed to the soil grains and a positively charged intermediate layer between the fixed layer and inner core Normally the positive and negative water layers adhere to each other but when an electric potential is applied between nearby electrodes the positive layer will flow towards the negative electrode dragging the uncharged water core with it By using standpipes for the negative electrodes the water can be collected in the standpipes and pumped out See CONSTRUCTION METHODS ELECTROOSMOSIS

[W H]

Electrophilic and nucleophilic reagent

Electrophilic reagents are chemical species which in the course of chemical reactions acquire electrons or a share [illegible]
or ions A [illegible]
dizing age [illegible]
agents are [illegible] thought of as cationic species such as H [illegible] NO_2 Br or SO_3 (or carriers of these species such as [illegible]
can [illegible]
Elec [illegible]
char [illegible] (cations) See ACID AND BASE

Nucleophilic reagents are the opposite of electrophilic reagents Nucleophilic reagents give up electrons or a share in electrons to other molecules or ions in the course of chemical reactions Nucleophilic reagents frequently are negatively charged ions (anions) Typical nucleophilic reagents are hydroxide ion (OH) halide ions (F Cl, Br, and I) cyanide ion (CN) ammonia (NH_3) amines alkoxide ions (such as CH_3O^-) and mercaptide ions (such as C_6H_5S) See SUBSTITUTION REACTION [J F B]

Electrophoresis

An electrochemical process in which colloidal particles are made to migrate under the influence of an electric current Particles of colloidal size dispersed in water are prevented from sticking together and thus separate out as a coagulum because of the repelling action of like electric charges These charges are borne on the particle surface and can result either from adsorbed ions that have been taken from the surrounding water or from charged atoms or groups of atoms that are an integral part of the chemical structure of the particle By virtue of these surface charges a colloidal particle will move toward an electrode of opposite charge just as do the ions of an electrolyte in solution during electrolysis (see ELECTROLYTIC CONDUCTANCE) The distinction between electrophoresis and electrical conduction by ions in solution is merely one of degree rather than of kind the colloidal particle is much larger than an ion and at the same time the colloid particle holds many more electrical charges than the 1 2 or 3 of single ions These two factors affect electrophoretic mobility in opposite ways The former decreases and the latter increases it so that ultimately the mobility of the colloid particle is not much different from that of ions namely between 2×10^{-4} and 20×10^{-4} cm/(sec) (volt) (cm)

Theory Suppose a nonconducting spherical particle of radius r cm immersed in a fluid of viscosity η poises and dielectric constant D bears a net charge of Q coulombs under the influence of a potential gradient x volts/cm let the particle move with a velocity v cm/sec The electrical force producing migration given by $Qx \times 10^7$ dynes is equilibrated by viscous frictional resistance which by Stokes' law equals $6\pi\eta rv$ Introducing the electrophoretic mobility $u = v/x$ and rearranging gives $u = (Q \times 10^7)/6\pi\eta r$ By this equation it is possible in theory to calculate the net charge Q on a particle from measurements of its electrophoretic mobility In many applications however the equation has been modified to meet the following criticisms (1) for nonspherical particles Stokes law cannot be applied (2) the frictional resistance encountered by the particle should also include a term for the retardation produced by the countermovement in the electric field of oppositely charged ions which because they are always solvated cause a net flow of solvent in the direction opposite to that taken by the particle

Measurement Apparatuses for measuring electrophoretic mobilities are of two sorts depending on whether the particles are microscopic or submicroscopic For microscopic particles (diameters of 1 10μ) the actual movement of an individual particle can be observed directly under the microscope its velocity is measured by timing its move-

ment across the plane of a calibrated graticule placed in the microscope eyepiece The glass microelectrophoresis cell used for this purpose has the form of a hollow microscope slide to each end of which the electrodes are attached Such an apparatus is suitable for measuring the electrophoresis of metal sols emulsions suspensions or bubbles and of biological materials such as bacteria blood cells or fungi The microscopic method can also be used by adsorbing colloid molecules for example soaps or proteins onto microscopic particles of glass quartz, or plastic The electrophoretic mobility that results is characteristic of the surface charge which may however have been modified by adsorption from that of the original hydrophilic colloid

A more satisfactory measurement of the electrophoresis of soluble colloids is to be had from the moving boundary method which when the boundary is observed by use of a cylindrical (schlieren) lens, enables refractive index gradients as small as 1 part in 6 000 000 to be detected and recorded photographically Important recent advances in the use of this method were made by Arne Tiselius The Tiselius electrophoresis cell is shown in Fig 1 The compartment is filled with the buffer solution of protein to a level just above α the center of the cell is then slid to the left the upper compartments emptied rinsed with buffer and then filled

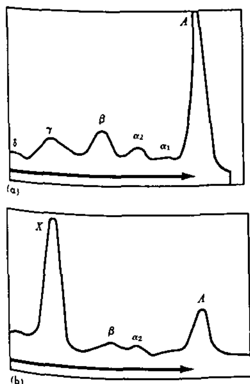


Fig 1 Diagrammatic representation of the Tiselius electrophoresis cell showing (a) formation and (b) motion of electrophoretic boundaries of a mixture of proteins (From G A Baitsell ed Science in Progress Ser 2 Yale University Press 1940)

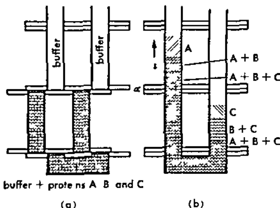


Fig 2 (a) Electrophoretic diagram of normal human blood serum showing mobility of albumin A and different globulins (b) Electrophoretic diagram of multiple myeloma blood serum The abnormal component labelled X migrates with γ-globulin A represents the albumin peak (American Instrument Co Inc)

with the buffer solution When the center section is then slid into place again a sharp electrophoretic boundary is formed On passage of the current i one side of the boundary moves upward and the other side downward The apparatus has evolved into a commercially available laboratory instrument which despite its necessarily elaborate design is widely used for routine clinical determinations The instrument produces a photographic diagram the abscissas represent the positions in the cell and the ordinates represent the refractive index gradients which can be related to concentration gradients the areas under the curves are therefore proportional to the concentration of the various components The great advantage that the method offers is in the analysis of protein components in naturally occurring proteins for example preparations that had previously been considered pure such as crystalline egg albumin or serum albumin have been shown by this method to contain two components By this means also blood serums have been analyzed and well marked differences have been discovered between normal and pathological

[illegible] component [illegible] such [illegible] cir[illegible]rhosis and nephrosis also cause equally marked changes in the electrophoretic diagrams of the blood serum or of the urine *See* ISOELECTRIC POINT [S RO]

Bibliography A Abramson L S Moyer and M H Gorin *Electrophoresis of Proteins* 1942 H Mark and E J W Verwey *Advances in Colloid Science* vol 3 1950

Electrophysiology (heart)

The science of the mechanisms, spread, and interpretation of the electrical manifestations of the heartbeat.

Electrocardiography. This is the study of the voltage changes occurring within the body and at its surface and caused by electrical excitation of the heart. The final interpretation of the observations is based on electrophysiologic principles and on empirical correlations of clinical value. *See* ELECTROCARDIOGRAPHY.

Vectorcardiography. Vectorcardiography is a form of electrocardiographic interpretation based on the assumption that the electrical manifestation of the heart can be treated as three-dimensional vector quantities and the voltage changes recorded from the body surface as the sum of various vectorial components. *See* VECTORCARDIOGRAPHY.

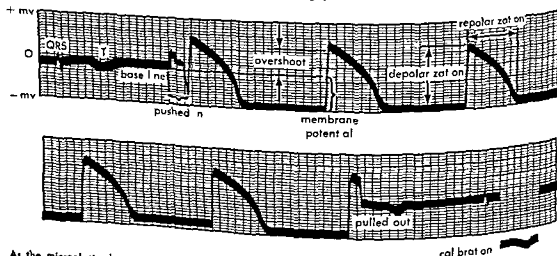
Heart cell potentials. Voltage differences between a single heart muscle fiber and its surroundings can be recorded in vitro and in vivo using glass capillary microelectrodes filled with 3 molar potassium chloride (KCl) solution. When the microelectrodes pass through a semipermeable ionic barrier (membrane) of a few molecules in width (100–200 angstroms) they permit a record of the cellular resting potential of 60–80 millivolts (mv). The cell interior is negative to the outside (see illustration). Concentrations of electrolytes differ on the two sides of the membrane. The known ratio of potassium ion (K^+) for heart muscle of 30K inside, 1K outside is sufficient to account for the observed potential difference at rest. This satisfies the Planck equations which relate voltage changes across semipermeable membranes to concentration differences. The cardiac excitation necessary for mechanical contraction is associated with a decrease in membrane resistance from 2000 to 50 ohm/cm² while the capacity remains unchanged at 2–10 microfarads/cm² ($\mu f/cm^2$).

The membrane becomes ion permeable, causing a current to flow with eventual reversal of the potential difference to about 30 mv, interior positive to outside (total cellular action potential: resting potential plus reversal, 90–120 mv). A sudden increase in permeability to sodium ion (Na^+) with a known ratio of $1Na^+$ inside to $10Na^+$ outside is at present considered the cause for the observed polarization reversal. During this potential cycle Na^+ enters the cell and thereafter K^+ leaves it. At the height of depolorization the process is reversed again. Na^+ is excreted by an active metabolic process known as the Na^+ pump and K^+ more gradually reenters the fiber. Restitution to the resting stage occurs slowly, depending on heart rate and species, and involves several metabolic phases.

Spread of excitation over heart muscle. Polarization reversal on the surface of a heart muscle fiber will cause it to become electrically negative with respect to an adjacent resting region of polarized fiber. In consequence, a current will flow from the resting to the active region which will change the polarization and initiate a response of the resting region. The resulting current, flowing from active to resting area, will cause firing of the previously resting cell. The resulting chainlike progression of changes in membrane polarization moving over the heart muscle is termed the cardiac action potential. The recorded potential variations at any point within a homogeneous volume of infinite extent are defined by Helmholtz' equation

$$V = \frac{M \cos \theta}{kr^2}$$

where θ is the angle subtended by the vector direction of the action potential with that to the recording point, r is the distance, and k is a constant.



As the microelectrode penetrates into the cell interior, a potential difference is recorded with the cell interior negative to the outside (membrane potential). On excitation the polarity is suddenly reversed (depolarization) and then gradually returns to the resting level (repolarization). The illustration shows a surface potential, insertion, and several cardiac cycles, and a pull out of the electrode. Calibration 50 mv; time lines 0.1 sec. (*From L. A. Woodbury, H. H. Hecht, and A. R. Christopherson, Am. J. Physiol., 164:307, 1951*)

characteristic of the medium M represents the dipole moment which is the charge times distance between resting and active areas and is dependent on the total mass activated (*see* DIPOLE MOMENT) Tissue inhomogeneities and boundary effects modify the final expression of such a surface electrogram Using cable equations the surface action *potential* of a single fiber may be considered the second derivative of its membrane current (*see* CALCULUS DIFFERENTIAL AND INTEGRAL) Mapping the *time course* of the total action current over the heart places the origin of cardiac excitation in the posterior *segment* of the right atrium the sinus venosus in which the normal pacemaker of the heart is located. Each individual heart muscle fiber however retains its inherent faculty for independent excitation and may at times act as a cardiac pacemaker *Radial* spreads of excitation over atrial muscle are contrasted with a more organized ventricular activation which occurs from within outward and is facilitated by a fast responding glycogen rich conducting system made up of modified muscle cells *known* as Purkinje fibers This system permits nearly instantaneous activation of the inner linings of the ventricular cavities of the heart *See* CARDIOVASCULAR SYSTEM HEART

Excitability of heart muscle Excitability *depends on* the presence of a resting potential although fibers may become unexcitable in the face of an adequate resting potential through such things as lack of Na^+, and blocking of membrane transfer by quinidine The refractory period during which the heart fails to react to any stimulus with a propagated electrical response coincides with the duration of cellular polarization reversal The period may be absolute during which all stimuli are ineffective or relative during which excitability gradually increases Depending on the nature of the test pulse the duration of absolute and refractory periods may vary Restitution to full cardiac excitability during the relative refractory period is also not a smoothly progressive event but shows fluctuations or dips of excitability Increased stimulus sensitivity may be found under certain circumstances immediately following the relative refractory period (the supernormal phase in conduction) The absolute duration of the excitatory state differs from fiber to fiber and under abnormal circumstances accentuated differences of this kind may cause the path of excitation to deviate from its normal course (aberrant conduction) These factors control many mechanisms responsible for irregular heart action *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY [H HE]

Bibliography C M Brooks B F Hoffman E E Suckling and O Orias *Excitability of the Heart* 1955 H H Hecht (ed) *The Electrophysiology of the Heart* Ann N Y Acad Sci vol 65 1957

Electroplating of metals

The application by electrodeposition of adherent metallic coatings to change the properties or dimensions of the surface of a metal Plating may improve the appearance hardness or resistance to

crease the

tective or

decorative but there is no sharp distinction Zinc and cadmium are applied to steel to prevent corrosion and their appearance is secondary Nickel and chromium are applied to automobile parts to preserve a good appearance but to do so they must also prevent corrosion Zinc and cadmium coatings protect small areas of steel that may be exposed through pores scratches or cut edges because under normal conditions these two metals are more readily attacked than steel They furnish electrolytic galvanic or sacrificial protection to the steel by forming a cell in which the zinc or cadmium is anodic and the steel is cathodic (*see* CORROSION ELECTROCHEMISTRY) Coatings of more noble metals such as copper nickel silver and gold do not prevent corrosion of exposed steel and may even accelerate it

In electroplating the cleaned article to be plated is connected as the cathode in a solution known as the electrolyte Direct current is introduced through the anode which usually consists of the metal to be deposited Metal dissolves from the anode and deposits on the cathode Under ideal conditions the same weight of metal dissolves from the anode as is deposited on the cathode and the over all composition of the bath remains constant These conditions are never fully realized and the bath composition changes and must be adjusted at intervals If the anode efficiency exceeds the cathode efficiency the metal content of the solution increases and the pH of the solution increases and vice versa In chromium plating insoluble anodes are used and metal must be added periodically to the solution by means of soluble compounds such as chromic acid

The constituents of a plating bath include a soluble compound of the metal to be deposited together with other substances added in fairly large amounts to increase conductivity throwing power or some other property Small concentrations of certain addition agents or brighteners are employed to yield smoother or brighter deposits Acid plating baths are cheaper to prepare and maintain but the alkaline baths which consist principally of complex cyanides have better throwing power and yield finer grained deposits

The throwing power represents the ability of a solution to produce coatings of uniform thickness on surfaces where the distances between various portions of the surface and the anode differ A cyanide copper bath may have a throwing power of +30% an acid copper bath of +5% and a chromium bath of −30% These numerical values are purely relative and depend upon the size and shape of the cell used The throwing power depends on (1) the rate of change of cathode polarization with current density (2) the bath conductivity and (3) the cathode current efficiencies at the prevailing current densities

The weight of metal deposited depends on the quantity of electricity (in coulombs or ampere hours) that is passed to the cathode The average thickness of the coating depends upon the current density (expressed in amperes per square decimeter or per square foot) and the period of deposition The uniformity of the coating thickness depends upon the shape of the article its position with respect to the anodes and the throwing power of the solution

The current density is an important factor in plating To produce a given current density it is necessary to apply a suitable potential expressed *in volts* The total bath potential includes (1) the decomposition potential (2) the *IR* drop which depends upon the resistivity of the bath and the distance between the anodes and cathodes (3) the cathode polarization and (4) the anode polarization at the prevailing current densities If as in most plating processes both the anode and the cathode deposit are of the same metal the decomposition potential is zero In early years the platers used only a voltmeter to control the operations but now they realize that ammeter readings are more significant Voltage readings may serve to indicate abnormal conditions *See* ELECTROLYSIS

METHODS AND APPLICATIONS

About 33 metals may be deposited from aqueous solutions while other *metals* can be deposited only from fused salt baths or organic electrolytes Those me[illegible]

[illegible] molybdenum and tungsten on the border line (*see* PERIODIC TABLE) The[illegible]

[illegible] metals are now plated commercially but the others can be so deposited if a demand arises for[illegible] to b[illegible] peri[illegible] [illegible] cop per silver and gold each form compounds with a valence of +1 such as copper(I) cyanide CuCN silver(I) cyanide AgCN and *gold(I)* cyanide AuCN which are employed in plating baths Copper also has a valence of +2 as in copper(II) sulfate $CuSO_4$ and gold has a valence of +3 in *gold(III)* chloride $AuCl_3$ All three of these are noble metals and hence easily deposited

Following are brief discussions of the metals commonly deposited with reference to their applications and methods of plating

Copper plating Copper is frequently applied to steel as an initial coating prior to nickel This usage has decreased since about 1935 except for those periods when the supplies of nickel were limited It has been found that under most service conditions composite coatings of copper plus nickel are inferior to those with the same total thickness of nickel only One advantage of copper under nickel is the greater ease of buffing the copper to a bright surface The present use of bright nickel deposits has largely eliminated this advantage of copper

One important use of copper plating is to prevent case hardening of steel on specified parts of the surface (*see* SURFACE HARDENING OF STEEL) For this purpose the entire article may be coated with copper which is then ground off from those areas on which hardening is desired Or the latter areas may be coated with an insulating varnish copper deposited on the remaining surface and the insulation dissolved off Copper is also deposited on steel or brass and treated with a soluble sulfide to produce an oxidized copper finish

Copper is deposited principally from acid sulfate and alkaline cyanide baths Baths containing sulfamate fluoborate or pyrophosphate are less extensively used The principal constituents of acid copper baths are copper sulfate and sulfuric acid with various addition agents to yield smoother brighter or harder deposits The cyanide baths contain cuprous cyanide and sodium or potassium cyanide In solution these combine to form complex cyanides such as sodium cuprocyanide $Na_2Cu(CN)_3$ An excess of cyanide known as free cyanide is always present and affects the anode and cathode efficiencies and the type of deposit In all cyanide baths carbonate is formed by decomposition of the cyanide and must be removed if its concentration becomes excessive

Silver plating The principal use of silver plating is on tableware because of its pleasing appearance and its resistance to attack by most foods (except those containing sulfur which tarnishes silver) The quality of silver plated ware is usually expressed by such terms as triple plate which represents a coating of 6 oz troy of silver per gross of teaspoons This corresponds to an average thickness of about 0 0008 in On the bowls of spoons and

Metals depositable from aqueous solutions (a part of the periodic table)

Group	VIa	VIIa	VIII			Ib	IIb	IIIb	IVb	Vb	VIb
Atom no Metal	24 Chromium	25 Manganese	26 Iron	27 Cobalt	28 Nickel	29 Copper	30 Zinc	31 Gallium	32 Germanium	33 Arsenic	34 Selenium
Atom no Metal	42 Molybdenum	43 Technetium	44 Ruthenium	45 Rhodium	46 Palladium	47 Silver	48 Cadmium	49 Indium	50 Tin	51 Antimony	52 Tellurium
Atom no Metal	74 Tungsten	75 Rhenium	76 Osmium	77 Iridium	78 Platinum	79 Gold	80 Mercury	81 Thallium	82 Lead	83 Bismuth	84 Polonium

forks an additional thickness of silver known as the overlay is usually applied

An important application of silver plating during World War II was on airplane *engine bearings for* which thick coatings were applied and followed by th n layers of lead and indium Silver is applied to the interior of waveguides and on electrical con tacts

Virtually all silver plating is done from solutions made by dissolving silver cyanide in a potassium cyanide solution Brighteners such as carbon disul fide are added to form bright deposits

Gold plating The most extensive application of gold plating is on jewelry where its pleasing yel low color and resistance to tarnish are desirable On cheap jewelry the gold may be as thin as 0 000002 in but on better grades up to 0 0001 in Gold is plated onto gold filled or rolled gold watch cases to cover the cut edges where brass would be exposed between two layers of the gold alloy Thick gold coatings are used on certain radar equipment and on electrical contacts

It is possible to deposit alloys of gold to produce desired colors, for example green gold with silver rose gold with copper and white gold with nickel Most of these colored gold coatings contain at least 85% gold corresponding to 20 carat gold (which contains 20/24 gold)

Practically all gold plating is done from dilute cyanide solutions because of the high cost of gold Very thin gold coatings are also produced by im mer

...rotect steel against corrosion It is difficult to apply zinc coatings thinner than about 0 002 in by hot dipping (see METAL COATINGS) For many ar ticles thinner coatings are adequate and are ap plied by plating known as electrogalvanizing Zinc plated coatings are usually purer than hot dipped coatings and have a slightly longer life for a given thickness On severe exposure zinc coatings tend to form bulky white corrosion products that may pre vent functioning of threaded parts To prevent this defect the plated articles are dipped into solutions that contain chromates to form conversion films that retard corrosion of the zinc

Zinc is deposited from both acid and cyanide baths The former are cheaper but have less throw ing power and hence are used for simple shapes such ... sed for a t ... prir

...de solutions contain zinc cyanide dissolved in sodium cyanide plus free cyanide and alkali which forms some sodium zincate

Cadmium plating Cadmium is much like zinc and also furnishes galvanic protection for steel In mildly corrosive or marine atmospheres cadmium coatings are about equal to zinc but in industrial locations the cadmium is inferior The high cost of cadmium confines its application to thin coatings on small parts such as threaded parts for airplanes fuses and other products It has less tendency than zinc to form white corrosion products

Cadmium is plated from solutions of cadmium cyanide in sodium cyanide plus some free cyanide and alkali Addition agents are used to form bright deposits

Tin plating The most extensive application of tin plating is in the production of electrolytic tin plate that is sheet steel coated with a thin layer of tin Hot dip tin plate usually has at least 1 2 lb of tin per base box (area covered by 112 plates each 14 by 20 in) One pound corresponds to an average thickness of 0 00006 in In World War II it became necessary to save tin By electroplating tin coat ings about 0 00003 in thick can be produced and have now largely replaced the hot dip tin plate Strip steel is plated at a rate of over 1000 ft/min and is then heated electrically or in oil to fuse the tin coating and make it less porous. Tin plating is also used on refrigerator coils and on small parts to be soldered

Acid tin baths contain stannous sulfate and sul furic acid and the alkaline baths contain sodium or potassium stannate Both types are used for strip plating but only the stannate for irregular shapes

Lead plating Lead coatings are used where re sistance to sulfuric acid is required for example on storage battery fittings The baths are acid and contain lead fluoborate or sulfamate with the cor responding acid Addition agents are used to yield fine grained deposits

Chromium plating Chromium is applied as the final layer on bright parts of automobiles and household equipment but is not the most significant part of the coating The chromium is only about 0 00002 in thick and is preceded by a thick layer of nickel under which may be a layer of copper Chromium is passive (not readily reactive) and th ...

Chromium plating is conducted from solutions containing chromic acid and sulfuric acid At each temperature bright deposits are obtainable within a fairly narrow range of current densities The cathode efficiency is only 10–15% and it increases with the current density causing the very poor throwing power of chromium baths For irregular shapes auxiliary anodes must be used to plate the surface completely

Nickel plating Nickel is far more extensively applied in electroplating than any other metal It is more resistant to atmospheric corrosion than other metals except noble metals such as gold and is fairly hard and ...

... is goal has not yet been fully realized During the shortage of nickel in World War II and the Korean War efforts were made to find substitutes for nickel or to produce satisfactory coatings with less nickel

Most nickel plating is done in baths containing nickel sulfate nickel chloride and boric acid Organic and inorganic additions are made to produce bright deposits There is no adequate explanation of the relation between the constitution and behavior of these brighteners By control of the bath composition temperature and current density it is possible to vary widely the physical properties of the nickel deposits

Alloy deposition It is possible to deposit two or more metals simultaneously to form alloy coatings with constitutions similar to cast alloys of the same composition The alloy most commonly deposited is brass an alloy of copper and zinc from cyanide solutions Its principal applications are for plating steel hardware and lighting fixtures to resemble brass and for plating steel to foster adhesion of rubber Most brass coatings contain about 80% of copper but resemble rolled brass with 70% copper Hundreds of other binary or ternary alloy deposits have been produced but few have found extensive applications

Preparation for plating To deposit adherent coatings it is necessary to have the surface of the basis metal clean that is free from all foreign substances such as grease and compounds such as oxides or sulfides The two essential steps are cleaning and pickling

Cleaning This is done to remove grease and attached solids Three principal methods are employed Solvent cleaning is done by vapor degreasing in which a solvent such as trichloroethylene is *boiled and its* vapors are condensed on the metal surfaces In ...

imm

kero

whic

troly... cleaning the articles are immersed in an alkaline solution and a direct current is passed between them and the other electrode Cleaning solutions may contain sodium hydroxide carbonate phosphate and metasilicate plus wetting agents and chelating agents More alkaline solutions are used for steel than for other metals Most of the cleaning is done by emulsification and not by saponification The articles may be connected as anodes as is usual for steel or as cathodes as is usual for other metals

Pickling This process removes oxides from the surface of the basis metal For steel sulfuric acid is used in large scale operations because it is cheaper but hydrochloric acid is also used for pickling because it acts faster Organic inhibitors are added to retard attack of the steel while the oxide is being dissolved *See* INHIBITOR (CHEMICAL) In cathodic pickling of steel attack of the metal is retarded while the oxide is being dissolved In anodic pickling in strong sulfuric acid a slightly etched surface is obtained

Hydrogen embrittlement is caused by absorption of hydrogen in steel during pickling and certain plating operations Especially with high carbon steels this causes a reduction in the fatigue strength The hydrogen is gradually evolved on standing and more rapidly by heating to about 200°C

Copper alloys are usually pickled in a solution containing sulfuric and chromic acids Brass is often given a bright dip in sulfuric and nitric acids

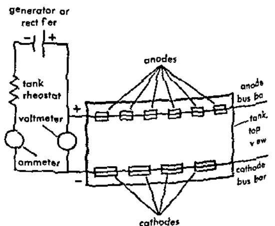
Electropolishing This is a process analogous to anodic acid pickling so conducted as to yield bright, smooth surfaces It has the following uses (1) to prepare surfaces for microscopic examination, (2) to produce a bright finish on metals such as stainless steel (3) as a taking off tool for changing dimensions, for example in removing steel from a bore of a gun barrel that is to be subsequently chromium plated to its original dimensions, (4) to prepare a surface for plating and (5) to brighten a plated surface

Most of the solutions for electropolishing steel contain sulfuric, phosphoric chromic or citric acids in suitable mixtures They are relatively concentrated and viscous and the products of the action are usually viscous or insoluble High current densities are used at elevated temperatures Any peaks on the surface are attacked more rapidly than the valleys in which viscous films tend to accumulate Hence the surface becomes smoother as measured by a profilometer

Equipment used in plating The important parts of typical electroplating equipment are shown in the diagram

Electrical equipment Most plating operations are conducted with direct current at potentials of 6–12 volts Since electricity is delivered as alternating current at 220 volts it is necessary to reduce the voltage and to rectify the current to direct current Both motor generators and rectifiers are used (*see* DIRECT CURRENT GENERATOR SEMICONDUCTOR RECTIFIER) Selenium rectifiers have been extensively used in plating but since 1955 germanium and silicon rectifiers have been introduced

Mechanical equipment The plating tanks are usually made of steel which requires no lining for alkaline solutions For neutral and acid solutions



Typical connections for a simple electroplating process

the tanks are lined with rubber or plastic For chromic acid baths lead linings are used and for nitric acid ceramic tanks are employed

Most large plating operations are conducted in conveyor tanks In semiautomatic conveyors the cathode racks are carried only through the plating tank In fully automatic conveyors the cathodes are carried successively through tanks that contain cleaning pickling and plating solutions with intermediate rinses This method saves hand labor and yields plated coatings that are more nearly uniform in thickness and quality

Small objects are plated in ' barrels usually hexagonal prisms with perforated plastic sides that rotate on a horizontal axis in a tank containing the plating solution and anodes The articles contact cathodic connections as they tumble during rotation of the barrel

SPECIFICATION AND TESTING

The increased use of electroplating especially in the automobile industry has led to the preparation of specifications by the American Society for Testing Materials and the American Electroplaters Society These include requirements and tests based on measurements and experience

Thickness The thickness of a plated coating is the most important factor in its protective value The minimum thickness is more important than the average thickness The latter may be determined in order to decide what weight of metal must be deposited to meet the required minimum thickness This is done by stripping the coating from a known area and determining the loss in weight of the sample or by analyzing the resultant solution Stripping may be done by immersing the sample in an appropriate solution or by making it anodic in a solution Similar methods may be used to remove defective coatings prior to replating The average thickness is computed from the weight of coating the area tested and the specific gravity of the coating The accuracy is limited by the uncertainty in measuring the area of irregularly shaped articles

The local thickness is measured at several points to locate the minimum Nondestructive methods are preferable because with destructive methods the cost of the articles destroyed may be greater than the cost of testing The nondestructive methods depend principally upon magnetic or electrical measurements

In a [illegible]
kn [illegible]
at [illegible]
It depends upon the force exerted by a spring that is required to detach a permanent magnet from the plated surface The instrument is calibrated by means of standard thickness specimens It can be used to measure nickel coatings on nonmagnetic metals such as brass nonmagnetic coatings such as copper on steel and magnetic coatings such as nickel on steel With other instruments the same principle is used but the magnet is detached by different means With certain instruments the magnetic reluctance is used to measure nonmagnetic metals on steel

With the Dermitron an electrical instrument also devised by A Brenner a high frequency current is applied and the eddy current serves as a measure of the thickness of nonmagnetic coatings on nonmagnetic basis metals provided the conductivity of the coating and that of the basis metal are different

In the microscopic method sections are cut perpendicular to the surface polished etched and examined under a microscope at a magnification of about 500 (*see* MICROSCOPE) Prior to sectioning copper may be deposited on the surface to yield a contrast and to protect the edges during polishing The thickness is measured with a filar micrometer or by projecting the image on a ground glass plate with a linear scale The microscopic method is often designated as the umpire method

Chemical methods depend upon the time required for a specified reagent to penetrate the coating In the spot test for chromium a drop of concentrated hydrochloric acid is placed on the surface and the period till bubbling ceases is measured In the dropping test for zinc and cadmium a reagent is dropped on the surface while in the jet test a continuous fine stream of liquid is run on the surface till the basis metal is exposed In the electrolytic test a known small area of the surface is made anodic in a suitable solution and a constant low current is passed until the voltage changes sharply when the coating is penetrated

Adhesion The adhesion of plated coatings is very important because poor adhesion may cause blistering or peeling of the coating and consequent corrosion of the basis metal Research methods have been developed to measure quantitatively the adhesion of coatings on specially prepared specimens but no method exists that is satisfactory for measuring the adhesion of coatings on plated articles Empirical tests such as bending twisting [illegible]

[illegible]

minimum thickness With more noble coatings such as copper nickel and chromium on steel the protective value depends principally upon the initial or subsequent presence of pores in the coating Present methods permit the detection of pores down to about 0 0001 in in diameter Smaller pores or other discontinuities may exist in the coatings In service pores may develop by enlargement of very minute pores or by attack of the coating at certain points as a result of concentration cells dissimilar metal contacts or foreign particles on the surface Studies are in progress on the cause and prevention of porosity of plated coatings

Accelerated corrosion tests To detect weaknesses in plated coatings accelerated tests have been used From about 1920–1950 the salt spray

test was extensively used The articles are exposed to a fine spray of fog of a neutral solution of sodium chloride until visible corrosion occurs Later experience has shown that this test does not yield reproducible or significant results and it has been largely abandoned The acetic acid salt spray test is more promising

Other methods under study include an electrolytic test in which the article is made anodic in an appropriate solution and the Corrodkote test in which the articles are coated with a slurry of clay and a salt solution and are then exposed to a high humidity for specified periods

Hardness and wear resistance Plated coatings may be used to protect metals against abrasion Wear resistance depends upon the hardness and other properties The hardness may be measured by conventional methods provided the depth of indentation is only a small fraction of the coating thickness (*see* HARDNESS SCALES) The Brinell hardness of plated metals ranges from 5 for lead to 1000 for chromium with a wide range for each met l

a u visibility The reflecting power depends upon *the reflectivity* of that metal and the specular reflectance of *the surface* which in turn depends upon the smoothness and such indefinable factors as haze Physiological and psychological factors may also affect the appearance *See* REFLECTION (ELECTROMAGNETIC RADIATION) Laboratory methods serve to measure the reflectance of plane surfaces but there is no present acceptable method of measuring the brightness of other plated surfaces Visual comparison with satisfactory standards is the usual basis for acceptance

SPECIAL APPLICATIONS

Electroforming This process produces or reproduces articles by electrodeposition A negative mold or matrix of the article is prepared by pressing or casting metal wax or plastic against the surface or by electrodepositing metal on the surface In special cases it is possible to produce a matrix or mandrel by machining stainless steel or aluminum To permit separation of the subsequently deposited replica of the original it is necessary to treat the surface of the matrix by apply ing films

surface is ... due conducting by the application of copper powder or of silver by chemical reduction or vacuum evaporation

The most important applications of electroforming are the production of electrotypes and phonograph matrices Minor uses include forming musical instruments venturi tubes fountain pen caps and models for research

Electrotyp

tic instead of wax as formerly Lead molds are used for fine halftones The plastic molds are coated with silver by chemical reduction lead molds are treated with a solution of a soluble chromate to permit separation of the deposit Most electrotypes are made by depositing a shell of about 0 00

bath For

deposited

This shell is separated from the mold and electrotype metal, an alloy of lead tin, and antimony is cast on the back of the shell

The plate thus made is finished by shaving it to specified thickness and making the printing face plane It is then mounted on a wood or metal base For cylinder presses the plate may be curved or cast curved For very long runs the printing surface may be given a layer of chromium

Plates for printing currency are made by electrodepositing the negative or alto a method that yields a more faithful reproduction The printing plates or bassos, are made by depositing a layer of nickel followed by a thick layer of iron The final printing surface is plated with chromium which gives a long life and can be dissolved off and replaced at intervals

An original phonograph recording is done in wax or plastic This is coated with a thin layer of gold by vacuum evaporation Copper is then deposited to form a master matrix On this are deposited several master records or mothers which serve as forms on which the pressing masters or stampers are deposited These may have an initial layer of nickel and a final layer of chromium to increase their life in the molding of the plastic records

Anodizing This is a process by which aluminum and magnesium are coated with a layer of oxide by making them anodic in an appropriate solution If as with magnesium alternating current is used it is certain that the oxide is formed during the anodic portion of the cycle though its properties may be modified in the cathodic portion As both aluminum and magnesium oxide are nearly insoluble and are good insulators it is surprising that a fairly thick coating can be built up The oxide forms at the metal oxide interface while some already formed oxide is being dissolved at the oxide solution interface The latter action produces fine pores into which the electrolyte penetrates and through which the current passes

Anodized aluminum Anodized aluminum is used for airplane assemblies and for ornamental purposes The process is conducted in either a chromic acid or sulfuric acid solution The former process is used for aircraft parts and the latter for many purposes that require a dyed surface In each method the hardness and porosity of the coating can be controlled by the concentration temperature and current density The protective value of the coatings can be improved by sealing which consists of treatment with hot water containing chromates or with live steam This hydrates part of the aluminum oxide and seals the pores Coat

ings made in sulfuric acid can be dyed with organic or inorganic compounds to produce colors, many of which are resistant to sunlight

Anodized magnesium These coatings consist principally of oxide with small contents of such compounds as phosphates, chromates, and chromium and manganese oxides, which color the coatings In the HAE process, developed at Frankford Arsenal in Philadelphia, the electrolyte contains potassium hydroxide aluminate, phosphate and manganate Alternating current is used at potentials up to 100 volts In the low voltage process, developed at the National Bureau of Standards the bath contains alkali and chromate Alternating current at 12 volts is used [W B]

Bibliography W Blum and G B Hogaboom *Principles of Electroplating and Electroforming*, 3d ed 1949, R M Burns and W W Bradley, *Protective Coatings for Metals*, 3d ed, ACS Monograph 129 1959, A K Graham (ed) *Electroplating Engineering Handbook*, 1955, A G Gray (ed), *Modern Electroplating*, 1953

Electropolishing

A step sometimes taken in the preparation of metals for electroplating, when an exceptionally smooth and bright surface is desired prior to the electrodeposition of the plating metal Electropolishing is accomplished by the anodic treatment of the base metal in any of several suitable solutions (most of them patented), and essentially combines the effects of both pickling and bright dipping The resultant surface is almost equal to that achieved by buffing, and the cost is much lower

In certain cases, as with some stainless steels and occasionally silver, electropolishing is the final step in the treatment of the metal surface and no hand buffing or plating follows *See* ELECTROPLATING OF METALS [G CO]

Electroscope

An instrument for detecting the presence and sign of an electric charge It is the simplest type of ionization chamber *See* IONIZATION CHAMBER

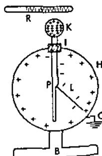


Fig 1 An electroscope being charged by induction by the negative charge on the hard rubber rod R

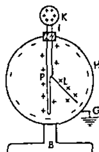


Fig 2 Positive charge left on the leaf of the electroscope after the induction process is complete

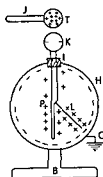


Fig 3 Testing the sign of an unknown charge on a test ball T

Figure 1 shows a common type of simple gold leaf electroscope Gold leaf, shown as L, is used because it is an extremely thin conducting foil which has low mass per unit area and is very flexible Hence it responds quickly and vigorously to small electrostatic forces Aluminum foil is almost as satisfactory as gold foil In Fig 1, P is a metal ends and windows so located that the motion and final position of L are visible H serves as a grounded electrostatic shield as well as a shield against air currents The base B supports the electroscope

The hard rubber rod R with its negative charge has set up the illustrated charge distribution by the process of electrostatic induction (*see* ELECTROSTATICS) The response shown is a test for the fact that R has a charge

To leave the electroscope with a net charge, a grounded conductor is touched to K so that the surplus electrons on P and L go off to ground, leaving the bound positive charge on K The ground connection is then broken and R is removed At this stage, shown in Fig 2, the electroscope is said to have a positive charge because there is a positive charge on its leaf system

If an electroscope has a charge of known sign, as in Fig 2 it can be used to test the sign of an unknown charge, as shown in Fig 3, where the metal test ball T, with its insulating handle J, has the un

known charge In the situation pictured L moves farther away from P as T is brought slowly up toward K showing that T has a positive charge If T had a negative charge, L would move toward P, as T slowly approaches K The converse situation, if the leaf system in Fig 3 were to have a negative charge initially can be readily visualized

Although electroscopes have been built with a wide variety of geometries, the principle of operation is essentially the same for all If an electroscope has a scale permitting quantitative measurements it is called an electrometer or electrostatic voltmeter For information on electrometers *see* VOLTMETER [R P WI]

Bibliography G P Harnwell *Principles of Electricity and Electromagnetism*, 2d ed 1949

Electrostatic lens

An electrostatic field with axial or plane symmetry which acts upon beams of charged particles of uniform velocity as glass lenses act on light beams The action of electrostatic fields with axial symmetry is analogous to that of spherical glass lenses whereas the action of electrostatic fields with plane symmetry is analogous to that of cylindrical glass lenses *Plane symmetry* as used here signifies that the electrostatic potential is constant along any normal to a family of parallel planes

The action of an electrostatic lens on the paths of charged particles passing through it is most readily visualized with the aid of an equipotential plot of the fields in a plane of symmetry of the lens The equipotential lines in the plot indicate the intersection with the plane of the drawing of surfaces on which the electrostatic potential is a constant The paths of charged particles in the electrostatic field are bent toward the normals of the equipotentials as the particles are accelerated and away from the normals as the particles are decelerated *See* ELECTRON MOTION IN VACUUM

Axially symmetric lenses These lenses are generally formed at or between circular apertures and cylinders maintained at suitable potentials A number of such lenses are shown with characteristic path plots in Fig 1 For any of these it is possible to define focal points principal planes, and focal lengths in the same manner as for light lenses and to determine with their aid image magnification for any object position (Fig 2) For a thin electrostatic lens in particular, that is a lens for which the extent of the variation in potential is small compared to its focal length, the object side focal length f_o and the image side focal length f_i are given by

$$\frac{\Phi_o^{1/2}}{f_o} = \frac{\Phi_i^{1/2}}{f_i} = \frac{3}{16}(\Phi_o\Phi_i)^{1/4}\int\left(\frac{\Phi'}{\Phi}\right)^2 dz$$

Here $\Phi(z)$ is the potential along the axis of the lens Φ' its derivative with respect to z (that is, the electric field along the axis), and Φ_o and Φ_i are the potential in object and image space, respectively The integration is extended over the

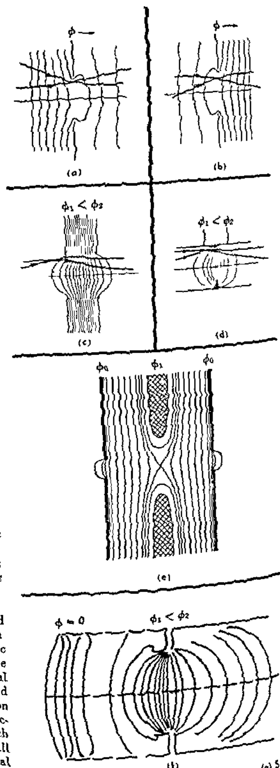


Fig 1 Axially symmetric electrostatic lenses (a) Single aperture lens (decreasing field) (b) Single-aperture lens (increasing field) (c) Two-aperture lens (d) Two-cylinder lens (e) Unipotential lens (f) Cathode lens (image tube) (From E G Ramberg and G A Morton, J Appl Phys, vol 10, 1939, and V K Zworykin, G A Morton, E G Ramberg, J Hillier and A W. Vance, *Electron Optics and the Electron Microscope*, Wiley, 1945)

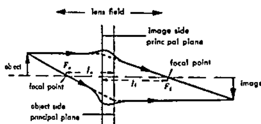


Fig 2 Definition of principal planes, focal points, and focal lengths for an axially symmetric electrostatic lens

lens field The quantity Φ is here normalized so that it is equal to the accelerating potential of the particle in question

Axially symmetric lenses are commonly divided into the four classes that follow

Simple aperture lenses These are the lens fields formed about circular apertures in a plane metallic electrode at potential ϕ with different electrostatic fields $-\phi'_o$ and $-\phi'_i$ on the two sides In most cases the focal length f of such a lens is given to a sufficient degree of accuracy by the Davisson Calbick formula for an aperture

$$\frac{1}{f} = \frac{\phi'_i - \phi'_o}{4\phi}$$

Simple aperture lenses are encountered as parts of more complex electrostatic lens systems as well as at the mesh openings of metal screens employed as electrostatic shields in vacuum tubes

Bipotential, or immersion, lenses In these lenses image space and object space are field free but at different potentials Typical examples are the lenses formed between apertures or cylinders at different potentials (Fig 1c and *d*) If the separation d of the two apertures is large compared to their diameters and if each component aperture lens satisfies the conditions for the validity of the Davisson Calbick formula, the focal lengths of the bipotential aperture lens are given by

$$\frac{1}{f_o} = \left(\frac{\phi_i}{\phi_o}\right)^{1/2}\frac{1}{f_i} = \frac{3}{8d}\left[1 - \left(\frac{\phi_o}{\phi_i}\right)^{1/2}\right]\left(\frac{\phi_i}{\phi_o} - 1\right)$$

The distances of the principal planes from the plane of symmetry are given by

$$h_o = -d/2 - 4d\phi_o/[3(\phi_i - \phi_o)]$$
$$h_i = d/2 - 4d\phi_i/[3(\phi_i - \phi_o)]$$

Quite generally, the principal planes are displaced from the plane of symmetry toward the low potential side with the image side principal plane closer to object space than the object side principal plane

For two cylinders of equal diameter D, whose difference in potential is small compared to their mean potential the focal lengths are given by

$$\frac{1}{f_o} = \left(\frac{\phi_i}{\phi_o}\right)^{1/2}\frac{1}{f_i} = \left(\frac{\phi_i}{\phi_o}\right)^{1/4} 0\,66\left(\frac{\phi_i - \phi_o}{\phi_i + \phi_o}\right)^2\frac{1}{D}$$

Bipotential lenses in particular lenses formed between two cylinders at different potentials, find wide application in beam focusing devices such as electron guns Like unipotential lenses, they invariably act as converging lenses

Unipotential lenses For this type the potentials are equal in object and image space In their simplest form these lenses consist of three apertures of which the outer two are at a common potential ϕ_o and the central aperture is at a different, generally lower potential ϕ_1 For such lenses with a central aperture of diameter D and the two outer apertures of smaller diameter separated a distance D from the plane of symmetry the weak lens focal length is given by

$$\frac{1}{f} = \frac{0\,2}{D}\left(\frac{\phi_o - \phi_1}{\phi_o}\right)^2$$

As ϕ_1 approaches zero the quantity $1/f$ increases more rapidly than indicated by this formula it attains a value of $0\,7/D$ for $\phi_1 = 0$ Unipotential lenses operated at high potentials (for example $\phi_o = 50$ kilovolts $\phi_1 = 0$) are employed as objectives and projection lenses in electrostatic electron microscopes (*see* MICROSCOPE, ELECTRON) The electrodes are commonly made out of stainless steel and given a high polish

Cathode lenses or immersion objectives Here the lens field extends from the emitter surface up to field free image space Examples are the cathode region of an electron gun the electron optical system of an electrostatic image tube or image converter and the objective of an emission electron microscope In the electron gun the cathode lens converges the electrons emitted by the cathode to a small spot the crossover, which is imaged by a second electron lens as the scanning spot on the cathode ray tube screen (*see* CATHODE RAY TUBE) In the image tube the cathode itself—a transparent photoemissive surface on which a light picture is projected—is imaged on a fluorescent screen beyond the cathode lens Frequently a cathode lens consists essentially of a uniform accelerating field followed by a short lens The image magnification m is then given by

$$m = v/2u$$

Here u is the distance between cathode and short lens and v is the distance between short lens and image The quantity v is given by

$$1/v = 1/f - 1/2u$$

where f is the focal length of the short lens

Lenses of plane symmetry. These lenses, analogous to cylindrical glass lenses are formed between parallel planes and at slits, replacing the circular cylinders and apertures of lenses with axial symmetry For the simple slit in an electrode at potential ϕ separating two regions of field $-\phi'_o$ and $-\phi'_i$ the focal length is given by the Davisson

Calbick formula for a slit

$$\frac{1}{f} = \frac{\phi'_s - \phi_o}{2\phi}$$

[E C R A]

Bibliography See ELECTRON MOTION IN VACUUM

Electrostatic precipitator

A device used to remove liquid droplets or solid particles from a gas in which they are suspended The process depends on two steps In the first step the suspension passes through an electric discharge (corona discharge) area where ionization of the gas occurs The ions produced collide with the sus pended particles and confer on them an electric charge The charged particles drift toward an elec trode of opposite sign and are deposited on the electrode where their electric charge is neutralized The phenomenon would be more correctly desig nated as electrodeposition from the gas phase The practical aspects of the electrostatic precipitator were demonstrated in 1906 by F G Cottrell

Construction In its simplest form the experi mental arrangement may consist of a vertical tube containing an insulated concentric wire (Fig 1) A d ...

pended particles are ionized in the corona dis-charge and migrate to the wall of the vertical tube If the suspended particles are liquid they will ac cumulate on the wall and coalesce into droplets which can be drained away from the bottom of the tube Solid particles may be displaced by mechani cal vibration or scrapers and discharged into a coni cal collector at the bottom of the tube

In more elaborate arrangements the ionization may occur in one vessel whereas the deposition occurs in a second stage A simplified two stage ap paratus is illustrated in Fig 2 In the first cham ber the particles become charged but are prevented

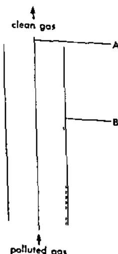


Fig 1 Diagram of a simple precipitator A corona wire B grounded tube

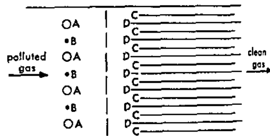


Fig 2 Diagram of a two stage precipitator vertical view A grounded cylinders B corona wires C grounded collector plates D charged plates of polarity similar to B

from depositing on the grounded cylinders by suit able adjustment of the rate of flow of the gas In the second chamber (consisting of alternately charged loosely packed parallel plates) precipi tation can be achieved satisfactorily by applying a lower potential than is necessary in the charging chamber since a corona discharge is not required

The corona discharge is usually produced by making the center wire negative because precipi tation efficiency is higher under these conditions However less ozone is produced by reversing the polarity and a positive wire is commonly employed in the cleaning of air when the presence of ozone may be objectionable The high voltage direct cur rent is commonly produced by mercury vapor or vacuum tube rectifiers The power requirements vary from 2–5 kw hr/1 000 000 ft^3 of gas being treated depending upon the amount and nature of the particles being removed Considerable difficulty is encountered in electrical leakage across the in sulators because drops of liquid or solid particles are deposited on them

H J White (1951) stated that suspended par ticles become charged by two different mechanisms The bombardment mechanism results from direct collisions between the suspended particles and ions produced in the corona discharge The diffusion mechanism consists of the attachment of ions to the suspended particles by ion diffusion Although both mechanisms operate simultaneously the former is more important However the latter effect may be come predominant for smaller particles perhaps in the 0 1 0 2 micron range

Application The use of electrostatic precipitators has become common in numerous industrial appli cations Each installation however is a separate design problem There is no comprehensive theory applicable in every case for the separation of par ticles or droplets from a moving gas stream which is usually in turbulent flow Consequently empir cal methods are generally used for designing pre cipitators In fact it is sometimes customary to make an accurate model of a proposed precipitator installation and to adjust and correct the model per formance in the laboratory to meet specifications prior to the manufacture of a scaled up unit

In view of the lack of a firm theoretical basis the efficiency of a given unit must be stated carefully

in reference to the particular unit with respect to operating voltage design geometry flow rate and particle removal The efficiency is usually expressed as the weight percentage of the material removed from the input stream Such an expression inadequately represents the effect of particle size It is usually desirable to remove the smallest particles from effluent gases since their higher degree of opacity makes their discharge more visible and therefore more objectionable to the public eye units are designed accordingly The large particles are therefore those which escape however the reverse condition may also be achieved by suitable design The efficiency is thus greatly influenced by the particle or droplet size distribution in the stream and an efficiency percentage based on weight must be *referred to the size distribution* most correctly before and after precipitation The efficiency depends exponentially on the stream residence time in the precipitator so that slower flow rates or longer vessels give better precipitation efficiencies

Among the advantages of the electrostatic precipitator are its ability to handle large volumes of gas at elevated temperatures if necessary with a reasonably small pressure drop and the removal of particles in the micron range Some of the usual applications are (1) removal of dirt from flue gases in steam plants (2) cleaning of air to remove fungi and bacteria in establishments producing antibiotics and other drugs and in operating rooms (3) cleaning of air in ventilation and air conditioning systems, (4) removal of oil mists in machine shops and acid mists in chemical process plants (5) cleaning of blast furnace gases (6) recovery of valuable materials such as oxides of copper lead and tin and (7) separation of rutile from zirconium sand *See* AIR POLLUTION CONTROL DUST AND MIST COLLECTION, PARTICLE PROPERTIES [CSM, WOM]

Electrostatics

The study of electric charges at rest under conditions where the positive and negative charges are separated from each other The term static electricity is often used to refer to electric charges at rest

Electrification Electrification by friction or rubbing occurs when one substance is rubbed with another and a surplus of electrons each with its negative charge is rubbed onto one of the bodies leaving a deficiency of electrons that is a positive charge on the other body The most commonly used materials for electrification by friction are hard rubber and fur The hard rubber has a negative charge after being rubbed with cat fur Only good insulators show a net charge after rubbing because electrical conduction permits neutralization of the charge very rapidly in other materials This effect is evident if the fur is held in the experimenter's hand during the rubbing process because electrical conduction through the experimenter's body will neutralize the positive charge nearly as fast as it appears on the fur One can get a net positive charge on a Lucite (Plexiglas) plate by rubbing it with a polyethylene plastic film The lucite will retain the charge because it is a good insulator *See* CHARGE ELECTRIC INSULATOR ELECTRIC

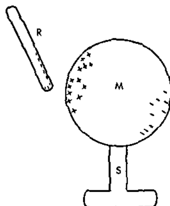


Fig 1 Charging by electrostatic induction The metal sphere M on an insulating stand S is charged by induction because of the presence of the negative charge on the hard rubber rod R

Electrostatic induction This process affords an other way of charging a body as illustrated in Fig 1 where the charged hard rubber rod R is responsible for the charging process Since like

M thus leaving an equal positive charge on the part of the sphere near R If a grounded lead (for example the experimenter s finger) is touched to the metal sphere the unbound electrons flow off to ground leaving the positive charge bound by the presence of the charge on R Then if the ground l nd next R removed the metal

static that is after it has completed its redistribution it is in any case all on the surface of the conductor

The induced charge is always equal in magnitude

objects Some of the induced charge is located on these nearby objects (the floor and walls and the experimenter s body)

Electrostatic generators These devices sometimes called static machines depend on electrification by friction and by induction for their operation The simplest electrostatic generator is the electrophorous shown in Fig 2 The hard rubber plate R has been given a negative charge by rubbing it with cat fur The metal plate D has an insulating handle H When D is set on R it touches it at only a few high points because R is micro

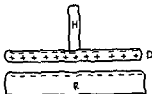


Fig 2 Electrophorous When the metal plate D with insulating handle H is placed on the rubber plate R charge is induced as shown

scopically rough even though it appears polished to the eye A positive charge is induced on the lower surface of D and an equal negative charge appears on the upper surface of D in the induction process Now if a grounded conductor is touched to D the electrons go off to ground and the ground connection is then removed Next by using the insulating handle H D is lifted from the rubber plate and there is a net positive charge on D which can be shared with a receiver and used for experimentation

Since no charge is removed from R while charging by induction D can be recharged from R as many times as the above process is repeated and a large charge can be built up on the receiver A rotating machine which would carry out the repetition of the induction process might be built from the electrophorous as follows In Fig 2 imagine an insulating wheel with a series of metal disks like D mounted on its rim and insulated from each other As the wheel rotates each disk in turn comes to the position of D in Fig 2 where a positive charge is induced on its lower surface and a stationary ground wire removes the equal negative charge from its upper surface When D moves away from its position over R the stationary ground wire loses contact and D leaves with its net positive charge which it delivers to the receiver at some other place along the path of its rotation As D moves away another metal disk comes into position over R and the process is repeated Thus charge will be transferred to the receiver as long as the wheel is turned For this purpose R could just as well be a metal plate which has been given a negative (or positive) charge by conduction because this will serve as well as the hard rubber plate as long as D does not come close enough to establish electrical contact between D and R The Wimshurst machine is an electrostatic generator based on this method of operation For information on the important electrostatic devices used in nuclear physics *see* COCKCROFT WALTON ACCELERATOR IMPULSE GENERATOR, RESONANCE TRANSFORMER, VAN DE GRAAFF GENERATOR, *see also* PARTICLE ACCELERATOR

It should be pointed out that there is actually no motion of positive charges in metals during current flow The description as given here is a convenient fiction which has been in common usage since Benjamin Franklin's time In metals it is the motion of the free electrons which constitutes the electric current, however, motion of electrons in one direction is equivalent to motion of positive charge in the opposite direction Thus the preceding description is entirely valid as long as this equivalence is kept in mind

Conservation of charge. The law of conservation of electric charge, an empirical law of experience says that the total net electric charge in the universe remains constant As has been explained, when electrification is produced by friction or by induction there is merely a separation of equal quantities [illegible] charge is r[illegible] ergy is ch[illegible] electric charges as in the creation of an electron positron pair from a γ ray photon one particle (the electron) has a negative charge and the other particle (the positron) has an equal positive charge Thus no net charge has been created Conversely, when annihilation of charged particles occurs as in electron positron annihilation to produce a pair of γ ray photons equal quantities of positive and negative charge disappear, and thus no net charge has been destroyed

Method of electric images. This is one of the schemes for solving certain types of electrostatic problems easily by the substitution of a physically simple but mathematically equivalent situation The types of problems where it is useful are those in which fixed free charges (such as q of Fig. 3a) have set up an induced charge on the surface of a conductor (such as the grounded infinite metal plane AB of Fig 3a) and a complete description of the electrostatic field in the region of the charge and external to the conductor is desired (for example, the region to the right of AB in Fig 3a) The equivalent situation used is one in which the conductor with its complicated distribution of induced charges is removed and replaced by a simple distribution of one or more fictitious charges which will produce the same electric field in the re-

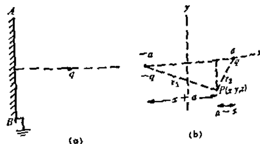


Fig 3 Illustration of the method of electric images. (a) Point charge q at a distance a in front of an infinite grounded conducting plane AB The problem is to find the potential function and equation for electric field intensity in the region to the right of AB the field being caused by point charge q and its equal distributed induced charge on the conducting plane (b) Physically simpler but mathematically equivalent situation for the region of interest

gion under study The fictitious charges which for the purpose of the solution replace the conductor with its induced charge are called the images of the fixed free charges

As an illustration the method of images will be used to solve the special problem posed in Fig 3a where q is a positive point charge considered to be on the x axis at a distance $x = a$ from the origin of coordinates located at the surface of the grounded infinite metal plane AB First however the problem in Fig 3b will be solved where there are two point charges $+q$ and $-q$ as shown and nothing else in the universe Using the principle of superposition for potential and the fact that electric potential is a scalar point function the potential at point P is given by

$$V = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{n} q_i/r_i$$

or in this problem

$$V = \frac{q}{4\pi\epsilon_0[(x-a)^2 + y^2 + z^2]^{1/2}} - \frac{q}{4\pi\epsilon_0[(x+a)^2 + y^2 + z^2]^{1/2}} \quad (1)$$

Here $\epsilon_0 \sim 8.85 \times 10^{-12}$ coulomb/newton m^2 is called the permittivity of empty space

From Eq (1) the equipotential surfaces can be determined by assigning specific values to V The particular equipotential surface for which $V = 0$ is the surface where $x = 0$ because the two terms on the right in Eq (1) then become equal in magnitude as well as opposite in sign Hence the yz plane is the equipotential surface for $V = 0$ and lies midway between the two point charges Thus the field would be the same if a grounded infinite plane conductor were placed in the yz plane and this is just the situation in Fig 3a the problem to be solved Therefore the field is the same to the right of the yz plane in Fig 3a and 3b and Eq (1) is the po en [illegible]

c [illegible]

th [illegible]

o [illegible] ue left of the yz plane is said to be the electric image of the charge on the right The gradient of the potential function gives the vector equation for the electric field **E** and with equations for **E** and V known most questions with regard to the physical problem can be answered Also the image force on q due to the distributed charge induced on the infinite plane can be seen to be given by Coulomb s law applied to the two point charges of Fig 3b *See* COULOMB S LAW *see also* CAPACITANCE DIELECTRICS ELECTRIC FIELD ELECTROSCOPE POTENTIAL ELECTRIC SHIELDING ELECTRICAL

[R P WI]

B bl ograph [illegible] *tric ty and* [illegible]
and N A [illegible]
1958 E F [illegible]
R P Win [illegible]

Electrostriction

A form of elastic deformation of a dielectric induced by an electric field specifically the term applies to those components of strain which are independent of reversal of the field direction Electrostriction is a property of all dielectrics and is thus distinguished from the converse piezoelectric effect a field induced strain which changes sign upon field reversal and which occurs only in piezoelectric materials *See* PIEZOELECTRICITY

Electrostrictive strain is approximately proportional to the electric susceptibility elastic compliance and the square of the field strength and is extremely small for most materials

The electrostrictive effect in certain ceramics is employed commercially in electromechanical transducers for sonic and ultrasonic applications *See* MICROPHONE TRANSDUCER UNDERWATER

[R D W]

Electrotherapy

Medical treatment which uses electrical techniques to change structure or function of the body or a body part In some of the methods the electricity is clearly not the major active agent in others the mechanism of action is unknown Therefore the classification is of questionable significance except as it indicates technique Only the following will be considered here High frequency currents are used within physiological limits as in diathermy or beyond such limits as in the destruction or removal of tissue by so called electrosurgery Another application is in closing small bleeding vessels in surgery electrocauterization Low frequency currents are used for passive exercise of paralyzed muscles Iontophoresis depends upon the movement of ions in a voltage gradient If two electrodes are in contact with tissue (such as skin and mucous membranes) then positively charged ions can be driven into the tissue by a positive electrode and negatively charged ions by a negative electrode Electroshock therapy is a method of producing violent excitation of the central nervous system which seems beneficial in certain types of psychopathology

High frequency currents In diathermy high frequency currents passed through the body between relatively large appropriately positioned electrodes generate heat just as does current flowing through any resistance The physiologically significant quantity is the temperature rise Low frequency or direct current also produces heat but the local polarization effects at the electrodes are destructive At frequencies used in medical diathermy any ionic changes produced at either electrode during one half a cycle are extremely small because the current flows for such a short time before it is reversed and such small changes as do occur are negated during the second half of the cycle when the electrodes change sign and the current changes direction Therefore the subject

experiences practically no sensation except that of heat and no muscle or nerve stimulation is produced A variety of apparatus types and any one of several wavelengths may be used Effects depend on duration and intensity of heating

The technique of diathermy is used to increase regional blood flow thus presumably improving local nutritional and oxygen supply as well as venous removal of metabolic end products As a result of the local heating more general actions on the body also occur The method can speed the resolution of subsiding local inflammatory changes is useful in treating subacute inflammations of bones, joints and bursae, and can be helpful in certain abdominal and thoracic pathology It can also be dangerously misused and is absolutely contraindicated in acute inflammations accompanied by suppuration and fever wherever there is tendency to hemorrhage and where there is malignancy

In electrosurgical methods the current flows between one small active electrode and a large indifferent one The high current density under the small electrode produces intense heat locally with tissue destruction Depending on current and on the apparatus design it is possible to achieve different degrees of destruction over controlled areas It is also possible to cut through tissue with an appropriately shaped electrode The advantages of electrosurgery stem from lessening of bleeding from small vessels facilitating rapid and gentle handling of fragile and very vascular tissues and lessening postoperative shock and pain

Low frequency currents These currents are below 10 000 cps, and unlike high frequency currents, flow long enough in one direction to produce local ion concentration changes in tissues and therefore can stimulate nerves and muscles They are used mainly for passively exercising muscles incapable of voluntary contraction This is thought to help retard atrophy and other degenerative changes in denervated muscle so that when the nerve regenerates it reconnects to a better preserved muscle machine

Iontophoresis In this technique the active electrode usually is a pad of absorbent material soaked with a solution of the material to be administered and placed in contact with the area into which the material should go, the indifferent electrode is a similar pad soaked with saline solution and placed elsewhere on the body Metal plates are placed on the pads and are connected to an appropriate source of direct current A positive electrode will drive positive ions into the underlying tissue, copper, mecholyl (methacholine), or zinc chloride can be so administered in certain skin gynecological upper respiratory, or eye pathology A negative electrode will inject negative ions such as iodide and chloride which are sometimes used in the treatment of skin scarring

Electroshock therapy. In electroshock therapy two electrodes are placed in appropriate positions on the head and 60-cps alternating current at about 70 volts is applied for about 0 1 sec The objective usually is to produce a convulsive seizure with immediate loss of consciousness and no memory of the experience, thus sparing the subject any unpleasant recall and making acceptance of subsequent treatments more certain There are modifications and varieties of this basic technique and there are also serious dangers and complications The method must be used only with properly selected patients and then with great skill and good judgement Properly applied it is a valuable therapeutic method in psychiatry *See* BIOPOTENTIALS AND ELECTROPHYSIOLOGY, PSYCHOTHERAPY

[J M T]

J D H Hock

Sh

mc

El

of Hydrotheraphy and Mechanotheraphy, 6th ed 1949

Elementary particle

Any of the theoretically irreducible constituents of the material world All particles of each kind are identical to one another This situation is inherent to quantum field theory, which describes particles as the quanta of fields, for each kind of elementary particle also called fundamental particle there is an associated field (*see* QUANTUM FIELD THEORY) The list of the known fundamental particles has not been and may never be, static With the advance of experimental techniques more particles are discovered, but with the advance of theory more particles can be understood as compounds of particles and hence not fundamental, for example atoms and atomic nuclei were at one time believed to be noncomposite

This article gives basic information on the elementary particles in tabular form and discusses their interactions and stability For detailed information on specific particles *see* ANTINEUTRON ANTIPROTON, BARYON, CASCADE PARTICLE, ELECTRON, GRAVITON, HYPERON, LEPTON, MESON NEUTRINO, NEUTRON, PHOTON, POSITRON PROTON, STRANGE PARTICLE, V PARTICLE For more complete discussions of topics mentioned see ANTIMATTER, ATOMIC STRUCTURE AND SPECTRA ELECTRON CAPTURE, ISOTOPIC SPIN, NUCLEAR STRUCTURE, PARITY (QUANTUM MECHANICS), QUANTUM ELECTRODYNAMICS, QUANTUM MECHANICS, QUANTUM STATISTICS QUANTUM THEORY NONRELATIVISTIC, RADIOACTIVITY, SELECTION RULES (PHYSICS), SPIN (QUANTUM MECHANICS) SYMMETRY LAWS (PHYSICS)

There are 31 particles currently (1960) regarded as elementary, these are listed in Table 1 in order of increasing mass The conventional symbol for each particle is given, the superscript is the charge of the particle in units of the electronic charge $|e|$, which is 4 8 × 10^{10} electrostatic units (esu) Of those written without a superscript the γ, ν, n, and Λ are neutral the p has charge $+|e|$ The symbol of the corresponding antiparticle

Table 1 Elementary particles

Name	T_3	Symbol	Rest mass Mev	Lifetime sec	Principal decay modes and their branching ratios	
'Classons' (massless bosons)						
Photon		$\gamma(\bar{\gamma} \equiv \gamma)$	0	∞		
Graviton			0	∞		
Leptons (weakly interacting fermions spin = $\frac{1}{2}\hbar$)						
Neutrino		$\nu(\bar{\nu})$	0	∞		
Electron		$e(e^+)$	0 51	∞		
Mu meson (muon)		$\mu^-(\mu^+)$	105 7	2 2 $\times 10^{-6}$	$e^- + \nu + \bar{\nu}$	100%
Sthenons (strongly interacting bosons spin = 0)						
Pi meson (pion), $s = 0$	1	$\pi^+(\overline{\pi^+} \equiv \pi^-)$	140	2 6 $\times 10^{-8}$	$\mu^+ + \nu$ $e^+ + \gamma$	$\cong$100% 001%
	0	$\pi^0(\overline{\pi^0} \equiv \pi^0)$	135	$<10^{-15}$ $>10^{-20}$	$\gamma + \gamma$ $\gamma + e^+ + e^-$	99% 1%
	-1	$\pi^-(\overline{\pi^-} \equiv \pi^+)$	See π^+			
K meson (kaon), $s = +1$	$\frac{1}{2}$	$K^+(\overline{K^+} \equiv K^-)$	494	1 2 $\times 10^{-8}$	$\mu^+ + \nu$ $\pi^+ + \pi^0$ $2\pi^+ + \pi^-$ $\pi^+ + 2\pi^0$ $\pi^0 + \mu^+ + \nu$ $\pi^0 + e^+ + \nu$	58% 25% 6% 2% 4% 5%
	$-\frac{1}{2}$	K^0 } { K_1 $(CP = +1)$	498	1×10^{-10}	$\pi^+ + \pi$ $2\pi^0$	$\cong$34%? $\cong$16%?
$s = -1$	$\frac{1}{2}$	$\overline{K^0}$ } { K_2 $(CP = -1)$	498	$\cong 10^{-7}$	$\pi^+ + \pi^- + \pi^0$ $3\pi^0$ $\pi^\pm + \mu^\mp + \bar{\nu}$ $\pi^\pm + e^\mp + \bar{\nu}$	} 50%
	$-\frac{1}{2}$	$K^-(\overline{K^-} \equiv K^+)$	See K^+			
Baryons (strongly interacting fermions spin = $\frac{1}{2}\hbar$)						
Nucleon $s = 0$						
Proton	$\frac{1}{2}$	$p(\bar{p})$	938 2	∞		
Neutron	$-\frac{1}{2}$	$n(\bar{n})$	939 5	$\cong 10^3$	$p + e^- + \bar{\nu}$	100%
Λ Hyperon $s = -1$	0	$\Lambda(\bar{\Lambda})$	1115	$\cong$2 6 $\times 10^{-10}$	$p + \pi^-$ $n + \pi^0$	$\cong$67% $\cong$33%
Σ Hyperon, $s = -1$	1	$\Sigma^+(\overline{\Sigma^+})$	$\cong$1190	0 8 $\times 10^{\ 10}$	$p + \pi^0$ $n + \pi^+$	$\cong$50% $\cong$50%
	0	$\Sigma^0(\overline{\Sigma^0})$	$\cong$1190	$<10^{\ 11}$	$\Lambda + \gamma$	$\cong$100%
	-1	$\Sigma\ (\overline{\Sigma}\ )$	$\cong$1196	1 7 $\times 10^{\ 10}$	$n + \pi$	$\cong$100%
Cascade hyperon, $s = -2$	$\frac{1}{2}$	$\Xi^-(\overline{\Xi^-})$	$\cong$1320	$\cong 10^{-9}$	$\Lambda + \pi$	$\cong$100%
	$-\frac{1}{2}$	$\Xi^0(\overline{\Xi^0})$	$\cong$1310	$\cong 10^{-9}$?	$\Lambda + \pi^0$	$\cong$100%

(charge conjugate particle) is written in parentheses The antiparticle has the same lifetime as the particle and its decay modes are the charge conjugates of those of the particle With fermions the antiparticle is always a distinct particle In the case of the neutral bosons, γ, graviton, and π^0 the particle and antiparticle are identical (self charge conjugate), these particles are thus eigenstates of charge conjugation, with the eigenvalues (charge parity C) -1, $+1$, and $+1$ respectively The neutral K mesons K^0, and $\overline{K^0}$ are each mixtures of equal parts of K_1 and K_2, which are eigenstates of CP (charge conjugation times space inversion)

According to quantum theory, the spin of a particle must be a half integral multiple of $\hbar$, where $\hbar$ is Planck's constant h divided by 2π If its spin is a half odd integral multiple of $\hbar$ it is a fermion, that is, it obeys the Fermi Dirac statistics, and behaves classically like a particle If its spin is an integral multiple of $\hbar$ it is a boson, obeying Bose Einstein statistics and behaving classically like a wave The massless, chargeless bosons (called classons in the table) are the quanta of familiar classical wave fields

Interactions. The primary interactions (couplings) of the elementary particles are of four types, as shown in Table 2 The letters F (fermion) and B (boson) represent the (possibly different) particles which are either emitted or absorbed at the point of the interaction Examples of the Yukawa interaction are $p \rightarrow n + \pi^+$, $K^+ \rightarrow p + \overline{\Sigma^0}$, of the four boson interaction $\pi^+ + \pi \rightarrow \pi^0 + \pi^0$, $K \rightarrow K^0 + \pi^- + \pi^0$, of the Fermi interaction $\bar{\nu} + p \rightarrow n + e^+$ Most of the foregoing interactions cannot satisfy energy and momentum conservation and hence occur only as virtual processes The coupling

Table 2 Primary interactions of the elementary particles

Type		Coupling strength	
		Value	Type
Yukawa	*FFB*	≅15 (pion)	Strong (nuclear)*
		≈2? (kaon)	Strong (nuclear)*
4 Boson	*BBBB*	?	Strong (nuclear)*
Current photon	J γ	≈1/137	Electromagnetic*
Fermi	*FFFF*	≈10 ²	Weak (beta)

* Renormalizable if fermions have spin ½ and bosons have spin 0

strength in the case of the current photon interaction (electromagnetic interaction) for instance is $e^2/\hbar c \sim 1/137 = \alpha$ the so called fine structure constant where e the electronic charge is the coupling constant of the interaction

Secondary interactions of particles arise from the combination of primary interactions For instance the emission of a photon from one charged particle and its absorption by another (called exchange of the photon) describes theoretically the observed electromagnetic interaction between the charged particles likewise exchange of π mesons between nucleons describes nuclear forces Other examples of secondary interactions are the scattering of light from an electron the scattering of light by light the decays $\pi^0 \rightarrow 2\gamma$ $\pi \rightarrow \mu + \nu$ and the interaction of a neutron with an electromagnetic field Secondary interactions are characterized by their finite range whereas primary interactions occur between particles at the same point Primary interactions themselves are secondarily modified by the combination of interactions (radiative corrections) These modifications are infinite but renormalizable except in the case of repetition of Fermi interactions

The elementary particles are grouped into weakly interacting particles (leptons and clasons) which have only weak and electromagnetic interactions and strongly interacting particles (baryons and sthenons) A striking feature of the strongly interacting particles is that they are grouped into charge (isotopic spin) multiplets the masses in each multiplet differ by only a few Mev The charge Q of a member of a multiplet depends linearly on the third component T_3 of the isotopic spin (i spin) $Q = |e (T_3 + n\ 2)$ where n is an integer describing the charge displacement of the multiplet Conventionally the term $n/2$ is written as $b\ 2 + s$ where b is the baryon number (+1 for a baryon −1 for an antibaryon 0 for a boson) and s is called the strangeness quantum number

Stability Except for the electron and proton all of the elementary particles having nonzero rest mass are unstable but it is considered strange that these unstable particles are as long lived as they are hence the name strange particle has been given to the heavier unstable particles (K mesons and hyperons) In the absence of any selection rule a strongly interacting particle would be expected to decay in about 10^{-23} sec their lifetimes are in fact of the order of 10^{-10} sec The explanation is that there is a selection rule In both strong and electromagnetic interactions charge and T_3 are conserved hence the charge displacement or equivalently the strangeness is conserved Inspection of Table 1 reveals that the charge displacements are such as to forbid all strong decays of the elementary particles

The decays of all elementary particles (except the electromagnetically allowed decays of π^0 and Σ^0) occur through the weak (generalized beta) interaction This interaction conserves lepton number and baryon number (hence the lightest baryon the proton is stable) it does not conserve T_3 (and therefore strangeness) parity or charge parity The nonconservation of strangeness apparently only extends to ±1 for the cascade hyperon does not decay directly to a nucleon plus a π meson which would involve a change of two units of strangeness [C J G]

Bibliography M Gell Mann and E P Rosenbaum Elementary particles *Sci American* 197 72 88 1957 M Gell Mann and A H Rosenfeld Hyperons and heavy mesons *Ann Rev Nuclear Sci* 7 407–478 1957

Elements (chemical)

An element is a substance made up of atoms with the same atomic number Examples of some common elements are oxygen hydrogen iron copper gold silver nitrogen chlorine and uranium ed Approx d the re sol ham (mercury and s ments

A few of the elements are found in na u in the free (uncombined) state Some of these are oxygen nitrogen the inert gases (helium neon argon krypton xenon radon) sulfur copper silver and gold Most of the elements in nature are combined with other elements in the form of compounds The most abundant element on the earth is oxygen the next most abundant is silicon The most abundant element in the universe is hydrogen and the next most abundant is helium See ELEMENTS (GEOCHEMICAL DISTRIBUTION)

The elements are classified in families or groups in a table called the periodic table Elements are also frequently classified as metals and nonmetals A metallic element is one whose atoms form positive ions in solution and a nonmetallic element one whose atoms form negative ions in solution See PERIODIC TABLE

While the atoms of a given element have the same atomic number they may not all have the same atomic weight (see ATOMIC WEIGHT) Atoms with identical atomic numbers but different atomic weights are called isotopes Oxygen for example is made up of atoms whose atomic weights are 16 17 and 18 Hydrogen is made up of isotopes 1 2 and 3 the isotopes of masses 2 and 3 are called deuterium and tritium respectively Carbon is made

Chemical elements including symbols, atomic numbers, and atomic weights

Name	Symbol	Atomic number	Atomic weight
Actinium	Ac	89	227 *
Aluminum	Al	13	26 98
Americium	Am	95	243 *
Antimony	Sb	51	121 76
Argon	Ar	18	39 944
Arsenic	As	33	74 92
Astatine	At	85	210 *
Barium	Ba	56	137 36
Berkelium	Bk	97	249 *
Beryllium	Be	4	9 013
Bismuth	Bi	83	208 99
Boron	B	5	10 82
Bromine	Br	35	79 916
Cadmium	Cd	48	112 41
Calcium	Ca	20	40 08
Californium	Cf	98	251 *
Carbon	C	6	12 011
Cerium	Ce	58	140 13
Cesium	Cs	55	132 91
Chlorine	Cl	17	35 457
Chromium	Cr	24	52 01
Cobalt	Co	27	58 94
Copper	Cu	29	63 54
Curium	Cm	96	247 *
Dysprosium	Dy	66	162 51
Einsteinium	Es	99	254 *
Erbium	Er	68	167 27
Europium	Eu	63	152 0
Fermium	Fm	100	253 *
Fluorine	F	9	19 00
Francium	Fr	87	223 *
Gadolinium	Gd	64	157 26
Gallium	Ga	31	69 72
Germanium	Ge	32	72 60
Gold	Au	79	197 0
Hafnium	Hf	72	178 50
Helium	He	2	4 003
Holmium	Ho	67	164 94
Hydrogen	H	1	1 0080
Indium	In	49	114 82
Iodine	I	53	126 91
Iridium	Ir	77	192 2
Iron	Fe	26	55 85
Krypton	Kr	36	83 80
Lanthanum	La	57	138 92
Lead	Pb	82	207 21
Lithium	Li	3	6 940
Lutetium	Lu	71	174 99
Magnesium	Mg	12	24 32
Manganese	Mn	25	54 94
Mendelevium	Md	101	256 *
Mercury	Hg	80	200 61
Molybdenum	Mo	42	95 95
Neodymium	Nd	60	144 27
Neon	Ne	10	20 183
Neptunium	Np	93	237 *
Nickel	Ni	28	58 71
Niobium	Nb	41	92 91
Nitrogen	N	7	14 008
Nobelium	No	102	253 *
Osmium	Os	76	190 2
Oxygen	O	8	16 000
Palladium	Pd	46	106 4
Phosphorus	P	15	30 975
Platinum	Pt	78	195 09
Plutonium	Pu	94	242 *
Polonium	Po	84	210
Potassium	K	19	39 100
Praseodymium	Pr	59	140 91
Promethium	Pm	61	147 *
Protactinium	Pa	91	231
Radium	Ra	88	226 05
Radon	Rn	86	222
Rhenium	Re	75	186 22
Rhodium	Rh	45	102 91
Rubidium	Rb	37	85 48
Ruthenium	Ru	44	101 1
Samarium	Sm	62	150 35
Scandium	Sc	21	44 96
Selenium	Se	34	78 96
Silicon	Si	14	28 09
Silver	Ag	47	107 880
Sodium	Na	11	22 991
Strontium	Sr	38	87 63
Sulfur	S	16	32 066
Tantalum	Ta	73	180 95
Technetium	Tc	43	99 *
Tellurium	Te	52	127 61
Terbium	Tb	65	158 93
Thallium	Tl	81	204 39
Thorium	Th	90	232 05
Thulium	Tm	69	168 94
Tin	Sn	50	118 70
Titanium	Ti	22	47 90
Tungsten	W	74	183 86
Uranium	U	92	238 07
Vanadium	V	23	50 95
Xenon	Xe	54	131 30
Ytterbium	Yb	70	173 04
Yttrium	Y	39	88 91
Zinc	Zn	30	65 38
Zirconium	Zr	40	91 22

* The atomic weights marked with asterisks are those of radioactive elements whose atomic weights depend upon the method of manufacture The listed isotope may be either the one of longest known half life or a better known one

The atomic weights are taken from the report of the International Commission on Atomic Weights published in the *Journal of the American Chemical Society* August 20 1958

up of isotopes 11, 12 13, and 14 Carbon 14 is radioactive and is used as a tracer in many chemical experiments All the elements have isotopes although in certain cases only synthetic isotopes are known Thus fluorine exists in nature as F^{19} but the artificial radioactive isotope F^{18} can be prepared Many of the isotopes of the different elements are unstable or radioactive and hence disintegrate to form stable atoms either of that element or of some other element See RADIOACTIVITY

Origin and uses of the elements The origin of the chemical elements is believed to be the result of the synthesis by fusion processes at very high temperatures (in the order of 100 000 000°C and

higher) of the simple nuclear particles protons and neutrons first to heavier atomic nuclei such as those of helium and then on to the heavier and more complex nuclei of the heavier elements In fact the energy of the sun and the stars is derived primarily from the fusion of hydrogen nuclei and electrons to form helium nuclei It is believed that this element producing fusion process is occurring even today in the hot stars *See* ELEMENTS AND NUCLIDES (ORIGIN)

The elements form the raw materials for the great chemicals industry today Various metals are used for structural materials protective coatings ornamental devices jewelry and tableware Such nonmetals as chlorine bromine hydrogen sulfur nitrogen are important for the manufacture of many of the common chemicals of commerce Helium is used to inflate dirigibles neon is used to make neon light signs and radon is used as a source of radioactive rays for therapy

A number of elements found in only very slight traces or not at all in nature have been synthesized Those elements are technetium promethium astatine francium and all the elements with atomic number above 92 These elements have been synthesized by a variety of nuclear reactions that involve transmuting atoms of one element into atoms of another by bombarding that element with neutrons or fast moving particles (protons deuterons and α particles) which will change the atomic number to that of the new element Not only have these elements been synthesized but isotopes of all the other elements also have been synthesized Today the hope of the alchemist has been realized in that all the elements have been transmuted *See* ATOMIC STRUCTURE AND SPECTRA CHEMISTRY, ISOTOPE

[A B C]

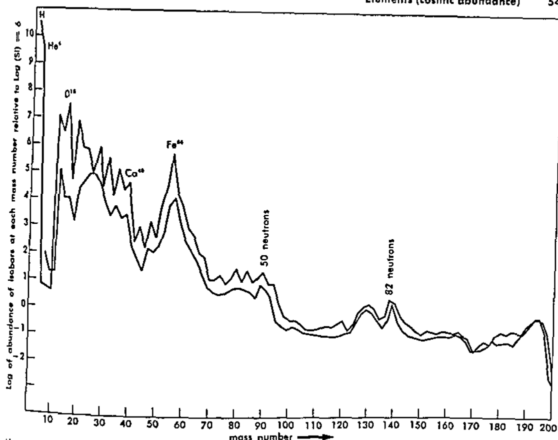
Elements (cosmic abundance)

The abundance of the elements in surface rocks of the earth in the earth as a whole in meteorites in our solar system in our galaxies or in the total universe corresponds to the average relative amounts of the chemical elements present or in other words to the average chemical composition of the respective object Element abundances are given in numbers of atoms of one element relative to a certain number of atoms of a reference element Silicon is commonly taken as the reference element in the study of the composition of the earth and the meteorites and the data are given in atoms per 10^6 atoms of silicon The results of astronomical determinations of the composition of the sun and of the stars are often expressed [illegible] of the composition of the rocks and meteorites The composition of the sun and of stars can be derived by quantitative spectral analysis On the surface of the earth the most abundant elements are oxygen silicon magnesium calcium, aluminum, and iron In the universe as a whole hydrogen and helium constitute more than 95% of the total matter *See* ACTIVATION ANALYSIS ASTRONOMICAL SPECTROSCOPY, ISOTOPE DILUTION TECHNIQUES

Abundances in the sun and stars The possibility of a quantitative spectral analysis of the sun stars, and planetary nebulae is based on the fact that the intensity of the absorption lines in the spectrum the so called Fraunhofer lines depends on the concentration of the atoms causing the absorption In order to calculate the relation of line intensity with atomic concentration a number of physical properties of the absorbing atoms as well as the thermodynamic state of the absorbing stellar matter have to be known in detail Furthermore a knowledge of the depth of the layer in which the absorption occurs the thermal velocity of the absorbing atoms their macroscopic turbulent motion and other characteristics is necessary before the exact functional dependence of line intensity and atomic concentration can be calculated

The first abundance data by spectral analysis were obtained by C H Payne Gaposchkin in 1925 and by H N Russell in 1928 Relatively few stars have been analyzed and the data from spectral analyses are far from complete However the data have shown that the chemical composition of the universe is remarkably uniform although systematic variations in the composition of stars seem to exist depending on age and position in the galaxies *See* SOLAR RADIATION, SUN

Abundances in the earth and meteorites The abundance of the elements in terrestrial rocks was first investigated by F W Clarke and H S Washington during the last decade of the nineteenth century These investigators compared numerous rock analyses and gave average figures for the occurrence of each element in the various types of terrestrial rocks These investigators had hoped that some sort of regularity might become apparent They expected that the chemical composition of the terrestrial rocks would reflect some fundamental quantity connected with the relative amounts in which the elements occur in nature in general Since then it has become obvious that meteorites are better objects for the study of a primeval abundance distribution of the elements The composition of the earth corresponds to the nonvolatile part of a primeval cloud from which the planets originated The meteorites have formed from the same cloud but they have undergone less chemical fractionation than any material on earth Meteoritic matter shows in general separation into three chemical phases metal sulfide and silicate in a ratio of about 10 6 1 100 respectively The elements that concentrate in the metal phase are called siderophile those that concentrate in the sulfide phase chalcophile and those in the silicate phase lithophile elements A large fraction of meteorites the chondrites contain all three phases in relatively constant proportions It is generally



Abundances of isobars plotted against their mass number The upper line refers to nuclei with even mass numbers the lower line to nuclei with odd mass numbers (After H E Suess and H C Urey)

believed that the chondrites contain the nonvolatile components of the primeval solar matter in essentially unchanged proportions because it seems improbable that chemically similar elements were separated from each other under conditions that did not lead to an effective separation of the three main phases from each other The giant planets (Jupiter, Saturn, and so on) have retained to a large degree volatile substances including hydrogen and helium, and elements such as carbon, nitrogen, and oxygen in the form of methane, ammonia, and water respectively *See* ATMOSPHERE, GEOCHEMISTRY OF, LITHOSPHERE, GEOCHEMISTRY OF, METEORITE

Nuclear abundances. Most elements are composed of more than one isotope (*see* ISOTOPE) The isoton [illegible] elements as a consequence of small differences in the chemical properties owing to the difference in mass Variations also occur if an isotope is produced by radioactive decay From the isotopic composition of an element and its cosmic abundance the nuclear abundances of its isotopes can be calculated A number of empirical rules exist for the abundances of nuclear species The most important one is Harkins' rule which states that isotopes with an odd mass number are less abundant than their even mass numbered neighbors In certain ranges it is possible to modify, within limits of error of the analytical data the values for the abundances of the elements in such a way that the abundance values of the individual nuclear species, as a function of their mass number, form regular smooth lines for the odd mass numbered species and at the same time, for the sum of the abundances of isobars at even mass numbers Irregularities occur where the number of neutrons or protons reaches a so called magic number, connected with a nuclear shell closure (*see* NUCLEAR STRUCTURE) Such modified abundance values as given by H E Suess and H C Urey are compared with the best empirical data on stellar spectra by L H Aller and on meteorites in the accompanying table

The nuclear abundances as a function of mass number are shown graphically in the illustration [illegible] line to those with odd mass number

Nuclear abundance values show a clear correlation with certain nuclear properties and therefore can be assumed to represent in good approximation the original yield distribution of the thermonuclear processes that led to the formation of the elements The empirical abundance values can therefore serve

Table of abundance values of elements

Element	Parts per million by weight			Atoms per 10^6 atoms of silicon	
	Meteorites*				
	Metal	Sulfide	Silicate	Astronomical values†	Cosmic abundances‡
1 H			630 ?	3 2 × 10^{10}	3 2 × 10^{10}
2 He				4 1 × 10^9	4 1 × 10^9
3 Li			3	0 23	100
4 Be			1	1 0	20
5 B			3		24
6 C	1100		400	11 × 10^6	11 × 10^6
7 N			1	3 × 10^6	3 × 10^6
8 O			4 × 10^5	3 1 × 10^6	3 1 × 10^6
9 F			40		1600
10 Ne				1 7 × 10^7	8 6 × 10^6
11 Na			7 8 × 10^3	6 2 × 10^4	4 4 × 10^4
12 Mg	320		1 6 × 10^5	4 8 × 10^6	9 1 × 10^5
13 Al	40		1 7 × 10^4	5 0 × 10^4	9 5 × 10^4
14 Si	40		2 1 × 10^5	1 × 10^6	1 × 10^6
15 P	2200	3100	1600	9 × 10^3	1 × 10^4
16 S	360	3 4 × 10^5	1 8 × 10^5	43 × 10^4	37 × 10^4
17 Cl			900	3 × 10^5	9 × 10^3
18 Ar				1 × 10^5	1 5 × 10^5
19 K			2000	2800	3200
20 Ca	500		2 × 10^4	74 000	49 000
21 Sc			5 8	49	28
22 Ti	100		1000	2900	2400
23 V	6		90	330	220
24 Cr	240	1200	3500	4900	7800
25 Mn	300	460	3000	6200	6900
26 Fe	9 1 × 10^5	6 1 × 10^5	1 6 × 10^6	1 8 × 10^5	6 0 × 10^5
27 Co	6300	100	200	1700	1800
28 Ni	8 6 × 10^4	1000	1400	2 9 × 10^4	2 7 × 10^4
29 Cu	310	4200	1 6	170	212
30 Zn	110	1500	3 4	1000	490
31 Ga	50	0 5	0 5	4 5	11 4
32 Ge	190	600	10	60	50
33 As	360	1000	20		4
34 Se	3	100	13		68
35 Br	1		25		13
36 Kr					51
37 Rb			4 5	5 2	6 5
38 Sr			26	19 5	19
39 Y			6 5	30	9
40 Zr	8		100	4 5	55
41 Nb	0 2		0 5	3	1 0
42 Mo	16	11	2 5	2 8	2 4
43 Tc					
44 Ru	10	4 2		8 8	1 5
45 Rh	4	1		11 5	0 2
46 Pd	3 7	4 5		0 3	0 7
47 Ag	3 3	21		0 01	0 3
48 Cd	8	30	1 6	1 8	0 9
49 In	1	0 8	0 2	0 2	0 1
50 Sn	80	15	3	1 0	1 3
51 Sb	2	8	0 1		0 25
52 Te		17			4 7
53 I	0 6		1 3		0 8
54 Xe					4 0
55 Cs			0 1		0 5
56 Ba			9 0	5 6	3 7
57 La			2 2		2 0
58 Ce			2 5		2 3
59 Pr			1 0		0 4
60 Nd			3 7		1 4
61 Pm					
62 Sm			1 3		0 7
63 Eu			0 3		0 19

Table of abundance values of elements (Cont)

Element	Parts per million by weight			Atoms per 10^6 atoms of silicon	
	Meteorites*				
	Metal	Sulfide	Silicate	Astronomical values†	Cosmic abundances‡
64 Gd			2.0		0.68
65 Tb			0.6		0.10
66 Dy			2.5		0.56
67 Ho			0.7		0.12
68 Er			2.1		0.32
69 Tm			0.4		0.03
70 Yb			2.0	1 ?	0.22
71 Lu			0.7		0.05
72 Hf			1		0.44
73 Ta	0.1		0.4		0.07
74 W	8.2		18 ?		0.5
75 Re	0.85				0.14
76 Os	7.6	10			1.0
77 Ir	3	0.5			0.82
78 Pt	19	30	0.1		1.63
79 Au	1.8	0.5	0		0.15
80 Hg		0.2 ?	0.01 ?	26 ?	0.02 ?
81 Tl		0.3 ?	0.15 ?		0.1 ?
82 Pb	60 ?	20 ?	2 ?	15 ?	0.1 ?
83 Bi	0.5 ?	2			0.01 ?
90 Th	0.04		2		0.03
92 U	0.01		0.4		0.018

* According to K Rankama and G Sahama
† According to L. H Aller
‡ According to H E Suess and H C Urey

as the basis for theoretical considerations about the origin of matter and of the universe These have led to the following conclusion no simple single mechanism exists by which the elements in their observed isotopic composition can have formed The matter surrounding us appears to be a mixture of material that formed under different conditions by different types of nuclear processes See ELEMENTS (GEOCHEMICAL DISTRIBUTION) ELEMENTS AND NUCLIDES (ORIGIN) [H E SU]

Bibliography S Fluegge (ed) *Handbuch der Physik* vol 51 1958 V M Goldschmidt *Geochemistry* 1954 K Rankama and T G Sahama *Geochemistry* 1950 H E Suess and H C Urey The abundance of the elements *Rev Modern Phys* 28 53–74 1956

Elements (geochemical distribution)

The earth has three major zones a high density core in part liquid a less dense solid mantle and a superficial crust (see LITHOSPHERE GEOCHEMISTRY OF) Of these three major zones only the crust which is 3 miles (oceanic) to 25 miles (continental) thick is available for direct sampling and chemical analysis and that only in part The composition of the mantle and the core are known from the study of stony and iron nickel meteorites respectively The degree of equivalence of the meteorite composition to earth interior composition is uncertain and data are accumulating which indicate that some differences exist The relative masses of material in each of the major geochemical divisions of the earth after B Mason *Principles of Geochemistry* 1958 are shown below

Geochemical divisions	*% of mass of earth*
Atmosphere	0.00009
Hydrosphere	0.024
Lithosphere	
Crust	0.7
Mantle	67.8
Core	31.5

Geochemical divisions of the earth The atmosphere contains a large portion of the noble gases of the earth and other compounds in the earth that have appreciable vapor pressure or are gases at surface temperature (see ATMOSPHERE GEOCHEMISTRY OF ATMOSPHERIC CHEMISTRY) The ocean represents another well defined geochemical phase See HYDROSPHERE GEOCHEMISTRY OF, SEA WATER

The composition of the core of the earth is generally taken as that of the iron phase of meteorites Likewise the first approximation composition of

cal purposes the crust can be conveniently divided into igneous rocks sedimentary rocks and deep sea sediments Another class metamorphic rocks may be adequately grouped chemically under either igneous or sedimentary rocks Thus a schist a metamorphic equivalent of a sedimentary shale has essentially the same major and minor elementary composition as the shale and a metamorphic gra

Distribution of elements in crust of earth, parts per million*

Element		Igneous rocks: Ultramafic rocks	Igneous rocks: Basaltic rocks	Igneous rocks: Granitic rocks: High calcium	Igneous rocks: Granitic rocks: Low calcium	Igneous rocks: Syenites	Sedimentary rocks: Shales	Sedimentary rocks: Sandstones	Sedimentary rocks: Carbonates	Deep-sea sediments: Carbonate cores	Deep-sea sediments: Clay cores	Crustal Model A
1 Hydrogen	H						See note A					
2 Helium	He						See note B					
3 Lithium	Li	0.X	17.0	24	40	28	66	15	5	5	57	29
4 Beryllium	Be	0.X	1	2	3	2	3	0.X	0.X	0.X	3	2
5 Boron	B	3	5	9	10	9	100	35	20	55	230	8
6 Carbon	C						See note A					
7 Nitrogen	N	6	20	20	20	20	See note A					20
8 Oxygen	O						See note A					
9 Fluorine	F	100	400	520	850	950	740	270	330	550	700	640
10 Neon	Ne						See note B					
11 Sodium	Na	4 200	8 300	28 400	25 800	40 400	9 600	3 300	400	20 000	40 000	24 200
12 Magnesium	Mg	204 000	38 000	9 400	1 600	5 800	15 000	7 000	47 000	4 000	21 000	11 900
13 Aluminum	Al	20 000	74 000	82 000	72 000	88 000	80 000	25 000	4 200	20 000	92 000	76 700
14 Silicon	Si	205 000	230 000	314 000	347 000	291 000	273 000	368 000	24 000	32 000	230 000	312 000
15 Phosphorus	P	220	1 000	920	600	800	370	190	400	350	60	800
16 Sulfur	S	2 500	2 000	1 000	500	500	2 600	300	1 100	1,300	1 300	970
17 Chlorine	Cl	40	60	140	200	280	180	10	150	30 000	30 000	150
18 Argon	Ar						See note B					
19 Potassium	K	2 040	6 800	25 200	42 000	48 000	26 600	10 700	2 700	2 500	25 000	29 200
20 Calcium	Ca	25 000	74 000	25 300	5 100	18 000	22 100	39 100	302 300	312 400	29 000	43 800
21 Scandium	Sc	70	30	14		3	14	1	1	3	25	14
22 Titanium	Ti	4 900	12 200	3 400	1 200	3 500	3 900	1 500	400	770	7 300	4 400
23 Vanadium	V	250	250	88	44	60	120	20	10	10	100	99
24 Chromium	Cr	1 600	170	22	26	2	87	7	7	11	93	65
25 Manganese	Mn	1 620	1 390	540	390	850	1 000	10	400	1 000	12 500	650
26 Iron	Fe	94 300	90 400	29 600	14 200	36 700	47 200	9 800	3 800	9 000	65 000	35 000
27 Cobalt	Co	110	48	7	1	1	19	2	1	8	100	12
28 Nickel	Ni	2 000	130	15	4.5	4	61	X	27	30	200	[illegible]
29 Copper	Cu	10	87	40	10	5	45	X	5	30	230	35
30 Zinc	Zn	150	112	47	39	26	43	16	20	10	50	55
31 Gallium	Ga	1.5	20	20	16.5	20	19	1	4	4	19	19
32 Germanium	Ge	0.5	1	1	1	1	1	0.5	0.1	0.2	2	1
33 Arsenic	As	1	2	2	2	2	13	1	1	1	13	2
34 Selenium	Se	0.42	0.33	0.17	0.08	0.08	0.6	0.05	0.08	0.17	0.17	[illegible]
35 Bromine	Br	1	3.6	4.5	1.3	2.7	4	1	7	100	100	[illegible]
36 Krypton	Kr						See note B					
37 Rubidium	Rb	7	30	110	170	110	140	60	3	10	110	119
38 Strontium	Sr	1	465	440	100	200	300	20	610	2 000	710	300
39 Yttrium	Y	0.X	21	35	40	70	26	40	5	15	150	34
40 Zirconium	Zr	45	110	140	175	500	160	220	18	18	180	160
41 Niobium	Nb	16	19	20	21	35	11	0.0X	0.3	4.6	14	14
42 Molybdenum	Mo	0.4	0.8	0.9	1.0	0.8	4.5	0.07	0.4	3	3	[illegible]
43 Technetium	Tc						See note C					
44 Ruthenium	Ru						See note D					
45 Rhodium	Rh						See note D					
46 Palladium	Pd	0.12	0.02	0.00X	0.00X		See note D					0
47 Silver	Ag	0.02	0.02	0.02	0.02	0.02	1.0	0.X	0.X	0.X	X	0
48 Cadmium	Cd	0.X	0.22	0.13	0.13	0.13	0.X	0.X	0.X	0.42	0.42	0
49 Indium	In	0.013	0.11	0.0X	0.26	0.0X	0.054	0.0X	0.0X	0.0X	0.08	0
50 Tin	Sn	0.5	1.5	1.5	3	3	11.0	0.X	0.X	0.X	1.5	[illegible]
51 Antimony	Sb	0.1	0.2	0.2	0.2	0.2	1.5	0.0X	0.2	0.15	1.0	[illegible]
52 Tellurium	Te						See note D					
53 Iodine	I	0.5	0.5	0.5	0.5	0.5	2.2	1.7	1.2	0.07	0.07	[illegible]
54 Xenon	Xe						See note B					
55 Cesium	Cs	0.X	1.1	2	4	0.6	8	0.X	0.X	0.4	6	[illegible]
56 Barium	Ba	0.4	333	420	810	1 600	550	50	10	210	2 100	611
57 Lanthanum	La	0.X	15	25	29	50	19	29	3.6	11	108	24
58 Cerium	Ce	0.X	48	81	92	161	59	92	11.5	35	345	79
59 Praseodymium	Pr	0.X	4.6	7.7	8.8	15	5.6	8.8	1.1	3.3	33	[illegible]
60 Neodymium	Nd	0.X	20	33	37	65	24	37	4.7	14	140	33
61 Promethium	Pm						See note C					
62 Samarium	Sm	0.X	5.3	8.8	10	18	6.4	10	1.3	3.8	38	[illegible]
63 Europium	Eu	0.X	84	1.4	1.6	2.8	1.1	1.6	0.20	0.6	6	[illegible]
64 Gadolinium	Gd	0.X	5.3	8.8	10	18	6.4	10	1.3	3.8	38	[illegible]
65 Terbium	Tb	0.X	84	1.4	1.6	2.8	1.0	1.6	0.20	0.6	6	[illegible]
66 Dysprosium	Dy	0.X	3.8	6.3	7.2	13	4.6	7.2	0.90	2.7	27	[illegible]
67 Holmium	Ho	0.X	1.1	1.8	2.0	3.5	1.2	2.0	0.25	0.75	7.5	[illegible]
68 Erbium	Er	0.X	2.1	3.5	4.0	7.0	2.5	4.0	0.50	1.5	15	[illegible]
69 Thulium	Tm	0.X	0.17	0.28	0.32	0.56	0.20	0.32	0.04	0.12	1.2	[illegible]
70 Ytterbium	Yb	0.X	2.1	3.5	4.0	7.0	2.6	4.0	0.50	1.5	15	[illegible]
71 Lutetium	Lu	0.X	0.6	1.1	1.2	2.1	0.7	1.2	0.15	0.45	4.5	[illegible]
72 Hafnium	Hf	0.6	1.6	2.3	3.9	11	3.7	5.1	0.40	0.41	4.1	[illegible]
73 Tantalum	Ta	1.0	1.1	2.0	4.2	2.1	0.8	0.0X	0.0X	0.0X	0.0X	[illegible]
74 Tungsten	W	0.X	0.5	1.3	2.0	1.9	1.8	1.6	0.X	0.X	X.0	[illegible]
75 Rhenium	Re						See note D					
76 Osmium	Os						See note D					
77 Iridium	Ir						See note D					
78 Platinum	Pt	0.02					See note D					
79 Gold	Au	0.006	0.002	0.00X	0.00X	0.00X	0.00X	0.00X	0.00X	0.00X	0.00X	[illegible]
80 Mercury	Hg	0.0X	0.08	0.06	0.06	0.0X	0.5	0.03	0.03	0.X	0.X	[illegible]
81 Thallium	Tl	0.06	0.13	0.43	3.1	1.4	1.0	0.82	0.0X	0.16	0.8	[illegible]
82 Lead	Pb	3	6	12	19	12	20	7	9	9	90	14
83 Bismuth	Bi		0.007		0.01							
84 Polonium	Po						See note F					
85 Astatine	At						See note F					
86 Radon	Rn						See note F					
87 Francium	Fr						See note F					
88 Radium	Ra						See note F					
89 Actinium	Ac						See note F					
90 Thorium	Th	0.004	4	4.4	13	13	12	1.7	1.7	13	9.4	[illegible]
91 Protactinium	Pa						See note E					
92 Uranium	U	0.001	1	1.4	3	3	3.7	0.45	2.2	1.3	1.3	[illegible]
93 Neptunium	Np						See note F					
94 Plutonium	Pu						See note F					

(See facing p[illegible])

k

a tic gneiss has the general chemical complexion of gran tic igneous rocks See METEORITE see also METAMORPHIC ROCKS

Igneous rocks The major igneous rock types encountered in the crust can be classified into (1) basaltic rocks and (2) granitic rocks which are further subdivided into high calcium granitic rocks (granodiorites diorites) and low calcium gran tic rocks In addition ultramafic rocks are present in orogenic regions and associated with strongly differentiated mafic bodie They are quantitatively unimportant in the crust as usually defined (that is the rocks above the Mohorovicic seismic discontinuity) though they may be important in the upper mantle Basaltic rocks occur pr marily as lava flows dikes and sills on the continents they constitute the major part of the oceanic volcanic islands and presumably are the important rocks composing the crust under the oceans Granitic rocks are the common rocks observed on the continents where not overlain by sed ments (for example the Canadian Shield and S... ute a sig Syenites pes found ma nly on the continents They differ from the gran tic clan in having higher aluminum and alkali content than the latter See IGNEOUS ROCKS

Sed mentary rocks These rocks are the result of the redeposition of the products of mechanical and chem cal degradation of pre existing rocks Graywackes arkoses and conglomerates represent sedimentary rocks with strong mechanically degraded components The components of sedimentary rocks wh ch are the result of chemical reconstitution are

*X 00 0 X etc nd
† M
39 F
rocks
h gh um gran tic rocks
No e A The
or
sp
el
se uta y rocks

Note B The rare gases occur in the atmosphere in the follow ng amounts (% vol) hel um 0 0002 neon 0 0018 a gon 0 93 krypton 0 0001 xenon 0 000008 Helium is produced by rad oactive decay of uran um and thorium but is also lost to

4 X 10 0

Note C The elements technetium and prometh um do not occur naturally in the earth s crust

Note D Data for these are missing or unrel able

Note E The following elements are generally present n rad oact ve en
u an
c urr
natu
uran um a consequence of neutron capture by

(1) clay (2) calcium and magnesium carbonate and (3) a resistate fraction primarily composed of quartz

When any one of these components is dominant the rock types are designated as shale carbonate rock and sandstone respectively In geochemical calculations these chemically reconstituted fractions are of particular significance hence data on them are presented in the table See SEDIMENTARY ROCKS

Deep sea sediments Sediments on the deep ocean bottoms are about 2000–3000 ft thick and hence represent an important component of the earth s crust Two major end members may be designated (1) a clay fraction often reddish in color because of oxidized iron and (2) a carbonate fraction composed of coccolith and Foraminifera tests See MARINE SEDIMENTS

Methods of analysis The methods of chemical analysis of geologic materials depend on the concentration of the elements sought Standard wet chemical techniques are adequate for major element determinations For the minor and trace elements more elaborate techniques may be called for These include emission spectrography spectrophotometry flame photometry x ray fluorescence mass spectrometric stable isotope dilution and neutron activation It is the concern of the analyst not only to devise a technique which gives reproducible results for trace quantities but to seek assurance of the accuracy of his determinations Because of blank corrections matrix effects and other difficulties this ideal is not always attained The development of isotope dilution and neutron activation methods have done much to put trace element analyses on an accurate basis as they are less subject to these effects See ISOTOPE DILUTION TECHNIQUES RADIOACTIVITY (APPLICATIONS)

Reliability of data The reliability of the data of the accompanying table of abundances is a function then of the degree of attainment of (1) systematic and sufficient sampling of the earth s crust and (2) accurate determinations of the concentrations of the elements in each sample Rarely are both of these criteria attained to an opti

actual analyses Where data are missing or obviously unreliable various methods of estimation were used

Crustal abundances The crustal abundances of the elements are usually calculated from the igneous rock data since they are the dominant components of the crust and presumably for most elements the chemistry of the sediments will be a reflection of this

Unfortunately only the surfaces of the continents and oceanic islands are available for sampling The vast ocean bottom is recognized as having primarily basaltic affinities (from seismic evidence and from rare boulders dredged from the sea floor

and presumed to be derived from this layer) Seismic and gravity data indicate that the deeper parts of the continental crust are probably more basic (contain less silicon and more magnesium and calcium) than the surficially exposed rocks Allowing for these uncertainties two models of the crust commonly used in estimating crustal abundances are considered The first is based primarily on the proportion of exposed igneous rocks on the surface of the continents (model A) the second model interprets the geophysical data to demand a crust composed primarily of equal parts of high calcium granitic rocks and basaltic rocks (model B) This latter model is the preferred one with present information [K K T]

Elements and nuclides (origin)

The origin of the elements must be linked with problems of cosmology and the nature and origin of the universe It is natural to attempt to explain the origin of the elements by a synthesis starting with one of the two fundamental building blocks of the atomic nucleus protons and neutrons *See* NUCLEAR STRUCTURE NUCLIDE

Two main kinds of theory have been advanced In the Initial Creation Theory the elements are built rapidly soon after the creation of the universe In the second the Continuous Production Theory elements are formed over a long time scale in the interior of the stars by processes that are existent today Evidence from both nuclear physics and astrophysics now supports the second theory

The basic data The relative abundances of the elements and isotopes in the sample of the universe which can be examined must form the starting point for either theory The differences in composition between the sun earth and meteorites can be explained through the loss of the lighter and more volatile substances by the less heavy bodies in the

ers of the nearby stars are similar to that of the sun

A schematic curve showing the relative cosmic abundances of the elements plotted logarithmically against atomic weight (A) is shown in Fig 1 Hydrogen the lightest is the

to flatten Superimposed on this general trend are a number of peaks and separate groupings of elements which give the

variants of this theory depending on the nuclear mechanism required Both assume a cosmology in which the universe (matter plus radiation) was created at a definite time in a very condensed state, and has been expanding ever since

One alternative is that the elements were built in conditions of thermodynamic equilibrium, with the

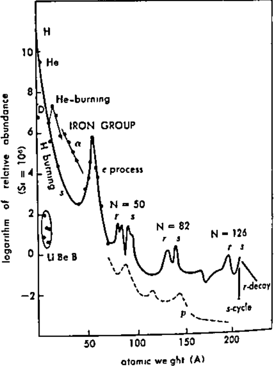


Fig 1 Schematic curve of atomic abundances as a function of atomic weight based on the data of H E Suess and H C Urey There is still considerable spread of the individual abundances about the curve illustrated but the general features shown are now fairly well established (*From Reviews of Modern Physics 29 547–650, 1957*)

temperature and density high enough for all nuclear reactions and their inverses to occur with great rapidity The rapid expansion of the universe at this early stage led to cooling and cessation of nuclear activity so that the relative abundances of the elements built were suddenly "frozen in" and should be both universal and the same as seen today In such a theory the abundances depend on the stability or binding energies of the nuclei However no single temperature and pressure can account for the observed relative abundances If the lighter elements are fitted reasonably well, far too few heavy elements are produced

The observed abundances of the heavy nuclei suggest that they have been mostly built out of neutron rich material Consequently two nonequilibrium theories were developed M Mayer and E Teller suggested that originally matter existed as a dense, "cold" nuclear fluid consisting essentially of neutrons This, being unstable split up in a way similar to the fission of heavy unstable elements such as uranium and the fragments formed the elements we have today This theory predicts the relative isotope ratios of the heavy elements fairly well but not of the light elements

G Gamow and his collaborators developed a nonequilibrium theory in which there is a primordial mixture of fundamental particles and radiation During the rapid early expansion of the universe

the protons (hydrogen nuclei) captured neutrons and eventually built all the elements right up to the heaviest This theory predicts the general features of the abundance distribution but not the details Its main problem however is that no stable nuclei exist at atomic weights 5 and 8 and it is not clear how the neutron capture chain could ever get past this point

Continuous production theory A theory based on the production of elements in the interior of stars avoids the objection against theories described above in that extreme conditions different from any thing known today need not be postulated The physics of stellar structure shows that the central temperatures and densities of stars are high enough for nuclear reactions to occur and that most of the energy of stars comes from such reactions

Since the first work on stellar origin theories by R d E Atkinson and Houtermans van Albada and principally F Hoyle more experimental nuclear physics data has become available theories of the changing structures of stars with time have become more secure and astrophysical observations have shown that the oldest groups of stars in our galaxy have a lower abundance of the heavier elements relative to hydrogen than the sun and younger stars A comprehensive theory of the origin of all the elements in stars starting from pure hydrogen was proposed in 1955–1957 by E M Burbidge G R Burbidge W A Fowler and F Hoyle Work has also been done along these lines by A G W Cameron It is found that eight separate processes occurring at different stages of the life history of stars are required to explain all the features of the abundance distribution

Energy comes from the building of heavier nuclei from lighter ones up to iron which has the greatest binding energy per nucleon of any element Lighter nuclei can react with one another at lower temperatures than heavier ones because of the greater electrostatic repulsion of nuclei with higher positive charges The greatest amount of energy is released in the first process the conversion of hydrogen to helium (hydrogen burning) During most of the life of a star mixing between its central core and outside does not occur but it has a self balancing mechanism so that when one nuclear fuel is exhausted in its core this contracts and heats until the next heavier fuel can be used

When the core has reached the stage of being composed of the iron group of elements no further energy supply is available Probably the star must become unstable and explode Stellar explosions called supernovae are seen in which most of the substance of the star disintegrates and spreads out into space

A theory of the origin of the elements in stars is not tied to a particular cosmology In any galaxy there will be a progressive enrichment in the heavier elements at the expense of hydrogen because the fuel exhausted stars explode and eject the elements built in their interiors back into the matter between the stars and new stars are continually condensing out of this enriched matter (although probably at a slower rate in our galaxy now than in its early history) Clusters of stars can be dated by astrophysical observations The oldest have less iron and less heavy elements than the sun by factors of 100–1000

Hydrogen burning In the first of the eight processes hydrogen is converted to helium and reactions between hydrogen and the light elements occur Starting from pure hydrogen successive captures of protons (p) form first deuterium (heavy hydrogen) and then the helium isotope of mass 3 two of which finally interact to produce He^4 If other substances are already present the final stages of this process (called the *pp* chain) can be modified or four protons can be combined to form He^4 by a catalytic cycle involving carbon nitrogen and oxygen (the main part of the cycle involving C and N was discovered independently by H A Bethe and C F von Weizsacker in 1938) In practice the *pp* chain occurs in lower mass stars with central temperatures less than 20 000 000° while the CNO cycle operates in more massive stars with higher central temperatures *See* CARBON NITROGEN CYCLE FUSION NUCLEAR PROTON PROTON CHAIN

Helium burning This process occurs when the core of a star has exhausted its hydrogen contracts and heats until helium nuclei (α particles) can interact at about 10^8 degrees Two α particles combine to form unstable Be^8 which at high enough temperatures and densities can capture a third α particle to form stable C^{12} before decaying This process was discussed theoretically by E E Salpeter in 1952 F Hoyle in 1954 and was experimentally confirmed by W A Fowler and others in 1957 Thus the difficulty experienced by Gamow s primordial theory where the density was much lower is avoided here This has led a school of Japanese physicists to attempt a theory in which the same conditions found in these stars occurred in the primordial matter but the cosmological details of such a theory have not been worked out

Further addition of α particles builds O^{16} and Ne^{20} Hydrogen and helium burning together can account for all the light nuclei up to neon in their

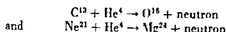
evidence for these processes hydrogen exhaustion high helium carbon and neon abundances abnormally high nitrogen/carbon ratio and a C^{12}/C^{13} isotopic ratio characteristic of equilibrium in the

core of a star contracts further and heats to about 10^9 degrees reactions can occur among the Ne^{20} nuclei built in helium burning freeing α particles These can be captured to build nuclei whose atomic weights are multiples of 4 up to Ca^{40} and possibly up to Ti^{48} This process accounts for the nuclei lying above their neighbors in the abundance curve labeled α in Fig 1

The e process When the temperature rises to about 3 000 000 000–4 000 000 000° all nuclear reactions occur in great profusion and thermodynamic equilibrium prevails The most stable nuclei (those grouped around Fe^{56}) with the greatest binding energies per nucleon will be built These elements rise in a high peak about a factor 10^3 above the general run of the abundance curve The equilibrium theory accounts very well for the relative abundances of all the nuclei in this region

The above four processes complete the energy giving reactions in stars

The s process This is the capture of neutrons produced slowly As first suggested by Cameron in 1954 when helium burning is occurring in a star conditions may arise in which C^{13} and Ne^{21} built previously in hydrogen burning react thus



$$C^{13} + He^4 \rightarrow O^{16} + \text{neutron}$$

and $$Ne^{21} + He^4 \rightarrow Mg^{24} + \text{neutron}$$

The neutrons will be captured predominantly by the iron group which have high capture probabilities and also by Ne^{22} A capture chain will result in building almost all the nuclei between $A = 23$ and 46 in particular those which were not built by the α process and many of the nuclei between $A = 63$ and 209 The time scale for neutron production and capture is slow enough that the necessary beta decays have time to occur between captures and building proceeds through stable nuclei Figure 2 shows how this building occurs The central stepwise line in a plot of the nuclear charge Z against atomic weight represents the stable nuclei and this is the path followed by the *s* process

This process builds high abundances of those heavy nuclei having closed shells of 50 82 or 126 neutrons which are particularly stable and will produce the peaks labeled *s* in Fig 1 at $A = 90$ 138 and 208 Isotopes of the elements strontium yttrium zirconium barium some rare earths and lead occur here and high abundances of these elements are found in certain stars The unstable element technetium observed in these stars but not found on earth is also built by the *s* process

The r process This involves the capture of rapidly produced neutrons At the end of a star's life the *e* process may touch off a catastrophic explosion during which an enormous flux of neutrons can be generated These will be captured predominantly by the iron group but in this case the flux will be so large that successive captures occur without time for intervening beta decays and the building chain will go through unstable nuclei on the right of the stability line in Fig 2 Abundances

Fig 2 Schematic plot showing how heavier nuclei are built from lighter ones by neutron capture on either a slow (s-process) or a rapid (r process) time scale The atomic number Z (charge of the nucleus) is plotted against the atomic weight A The build up goes from the bottom left hand corner upward and to the right Capture of a neutron moves a nucleus one unit to the right and beta-decay moves it one unit upward At the closed shells of neutrons ($N = A - Z = 50$ 82 and 126) addition of a further neutron is more difficult and excess abundances build up at these points (From Reviews of Modern Physics 29 547–650 1957)

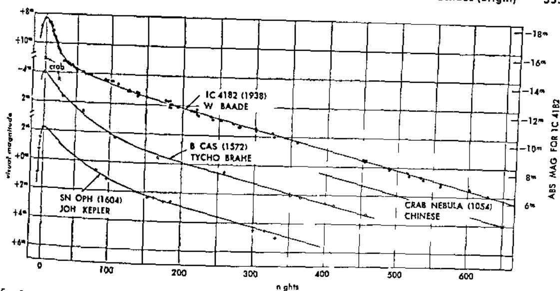


Fig 3 Light curves of supernovae by W Baade Measures for the supernovae in IC 4182 are by Baade those for B Cassiopeiae (1572) and SN Ophiuchi (1604) have been converted by him to the modern magnitude scale from the measures by Tycho Brahe and J Kepler The three points for the supernova of 1054 are uncertain, being taken from the ancient Chinese records The abscissa gives the number of nights after maximum the left hand ordinate gives the apparent magnitude on a logarithmic scale (separate scale for each curve) the points for the Crab Nebula belong on the middle scale—that is that for B Cassiopeiae The right hand ordinate gives the absolute magnitude for SN IC 4182 derived by using the current distance scale (From Reviews of Modern Physics 29 547–650 1957)

will pile up at the closed neutron shells in the unstable nuclei After the explosion ceases these unstable nuclei decay to stable ones and abundance peaks are produced slightly shifted from the s process peaks The peaks labeled r in Fig 1 at $A = 80$ 130 and 194 are thus accounted for

Supernovae show a rapid increase of light some times becoming about 1 000 000 000 or more times brighter than the sun They then fade apparently at a strictly exponential rate as shown by Fig 3 The only known explanation for fading of this sort is radioactive decay The unstable heavy element californium 254 which could be built in a rapid neutron capture process and beyond which building would probably cease decays by spontaneous fission with a half life of $56\ 2 \pm 0\ 7$ days and releases a large amount of energy The half life of fading of the best observed supernova is 55 ± 1 day This agreement suggests Cf^{254} is produced in a supernova of this sort and supports the theory of the r process which predicts production of Cf^{254} Fe^{59}, with a half life of 45 days is produced in an r process if fewer neutrons are available and has also been suggested as responsible for the decay curves of supernovae

The r process will build most of the remaining isotopes from $A = 70$ to uranium and those few between $A = 35$ and 50 not accounted for by other processes The calculated abundances between $A =$ 70 and 210 agree well with those observed

The p process The p process proton capture or ejection of neutrons by gamma rays is a modifying process necessary to account for the few heavy nuclei that cannot be built by either slow or rapid neutron capture chains These are all proton rich nuclei and are a factor of at least 100 less abundant than their neighbors (Fig 1) They can be produced by proton capture by nuclei already built by other processes either in supernova explosions of stars that still have considerable hydrogen left in their envelopes or in the outer parts of the supernovae thought to be responsible for the r process

The x process The light elements lithium beryllium and boron lie below the general abundance curve by a factor of the order of 10^7 They are all destroyed rapidly in hydrogen burning at quite low

However they can be formed by the splitting of heavier nuclei by fast charged particles under certain conditions of low density such as by cosmic rays in the gas between stars or in the atmosphere of stars with large magnetic fields in which a few protons may be magnetically accelerated to high energies

The abundance of deuterium 1/7000 of hydrogen on the earth is hard to account for quantitatively this way Deuterium has not been detected on the sun except in localized magnetic disturbances nor in the gas between stars Possibly deuterium on the earth was produced during the formation of the solar system Alternatively it might be produced under certain conditions in the expanding shells of supernovae Some may be produced in the surfaces of magnetic stars

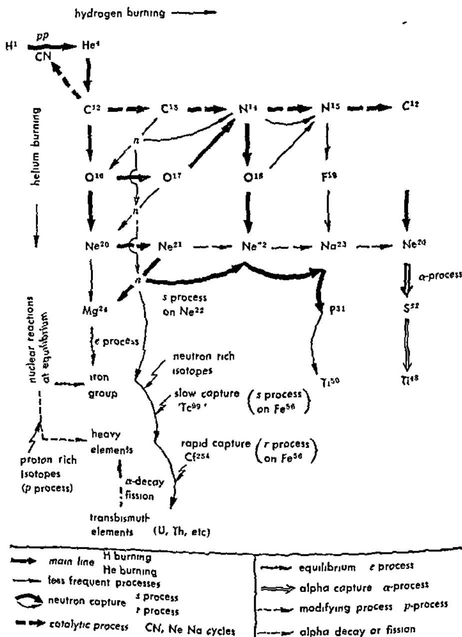


Fig 4 A schematic diagram of the nuclear processes by which the synthesis of the elements in stars is thought to take place Elements synthesized by interactions with protons (hydrogen-burning) are listed horizontally Elements synthesized by interactions with alpha particles (helium burning) and by other processes are listed vertically The details of the production of all of the known stable isotopes of carbon, nitrogen oxygen, fluorine, neon, and sodium are shown completely Neutron capture processes by which the highly charged heavy elements are synthesized are indicated by curved arrows (From Reviews of Modern Physics 29 547–650, 1957)

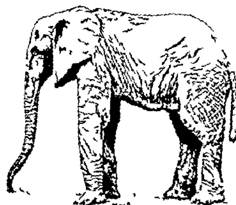
The groundwork has thus been laid for a complete theory of the origin of the element
Figure 4 sho
Improvemen
low with m
physics furt er calculations of structures of evolving stars and more determinations of the compositions of stars

[E M B]

Bibliography R A Alpher and R C Herman Theory of the origin and relative abundance distribution of the elements, *Revs Modern Phys* 22 153–212 1950, E M Burbidge G R Burbidge W A Fowler, and F Hoyle, Synthesis of the elements in stars *Revs Modern Phys*, 29 547–650 1957, E M Burbidge and G R Burbidge Formation of elements in the stars *Science* 128(3321) 387–399, 1958, A G W Cameron Nuclear reactions in stars and nucleogenesis *Publs Astron Soc Pacific*, 69 201–222, 1957, W A Fowler Formation of the elements *Sci Monthly* 84 84 1957

Elephant

Either of 2 living species of the order Proboscidea There are about 300 known fossil species The two surviving forms are the Indian elephant *Elephas indicus* and the African elephant *Loxodonta africanus* Although superficially similar, they have a radically different tooth pattern The African species is somewhat heavier and taller, weighing as much as 13,000 lb and having a height of 13 ft at



The African elephant *Loxodonta africanus* length to 10–12 ft (From E L Palmer *Fieldbook of Natural History* McGraw Hill 1949)

the shoulder Whereas both species are seen in circus tents and zoos the African species cannot be readily domesticated and the Asian species is usually trained to perform The Asian species is a valuable beast of burden and work animal being easily trained and handled Both species travel in large herds and may be highly destructive to plantations that lie in their path *See* PROBOSCIDEA

[J D B]

Eleutherozoa

One of the two subphyla of Echinodermata that include forms which are free living and not anchored to the sea bed All living echinoderms except crinoids are eleutherozoans The body is globular cylindrical or star shaped The mouth and anus lie at opposite ends of the body The mouth usually turns downward and is located in the middle of the lower surface In some forms there is no anus and the gut is a blind sac The water vascular system usually develops tube-feet with ampullae and these serve as locomotor organs Radial symmetry although sometimes inconspicuous is always present and usually pentamerous In the course of time some forms have departed from the typical pattern Thus some basket stars lie with the mouth turned upward in some heart urchins the anus has moved toward the mouth and most holothurians rest on one side

The ancestry of the Eleutherozoa may be found among the Edrioasteroidea There are three extant classes the Holothuroidea Echinoidea and Asterozoa as well as one extinct class the Ophiocystioidea. The oldest Eleutherozoa are from the Lower Ordovician *See* ECHINODERMATA EDRIOASTEROIDEA

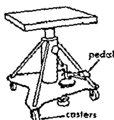
[H B F]

Elevating machines

Distinctive materials handling machines that lift and lower a load through a fixed vertical path of travel with intermittent motion In contrast with hoisting machines elevating machines support their loads instead of carrying them suspended and their paths of travel are both fixed and vertical They differ from vertical conveyors in operating with intermittent rather than with continuous motion Industrial lifts stackers and freight elevators are types of elevating machines

Industrial lifts A wide range of mechanically hydraulically and electrically powered machines are classified simply as lifts (Fig 1) They are adapted to such diverse operations as die-handling and feeding sheets bar stock or lumber In some locations where differences in floor level occur between adjacent buildings, broad platforms which serve as floor levelers obviate the need for ramps They are also used to raise and lower loads between the ground and the beds of carriers where no loading platform exists

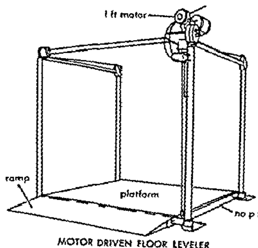
Lifting tailgates attached to the rear of trucks are useful for handling drums barrels crates, household appliances and similar articles particu



HYDRAULIC ELEVATING WORK TABLE



HYDRAULIC LIFT FLOOR LEVELER



MOTOR DRIVEN FLOOR LEVELER

Fig 1 Industrial lifts

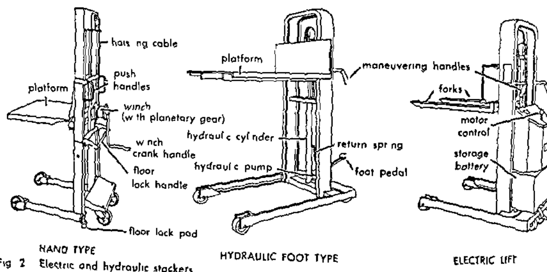


Fig 2 Electric and hydraulic stackers

larly at delivery points where no unloading facilities are provided They are usually actuated by battery operated motors or by a power takeoff from the drive transmission

Advances in mechanized handling made necessary sturdier more efficient dockb... plates As a ... operated loa... of these ad... tors ...

with a vertical frame which supports and guides a carriage to which in turn is attached a platform pair of forks or other load carrier (Fig 2) The carriage may be raised and lowered by hand by an electrically driven winch or by a hydraulic cylinder which actuates a system of chains or cables and which may be lever pedal or power operated Earlier electric motors used on stackers were plugged into power lines to receive current but the trend is to make the machines independ... and today th... teries or b... stackers are ... usually have ... during operation

Basic types of stackers can be varied in several particulars The masts can be hinged or telescopic and the platforms may be plain equipped with rollers or constructed especially to handle specific products Some stackers have devices for tilting barrels and drums or for lifting and dumping free flowing bulk materials

Stackers have a significant place in the development of materials handling equipment They are the prototypes of completely powered noncounter balanced platform and forklift trucks See INDUSTRIAL TRUCKS

Industrial elevators Examples of industrial elevators range from those set up temporarily on construction jobs for moving materials and personnel between floors to permanent installations for mechanized handling in factories and warehouses. Electric dumbwaiters with capacities up to 500 lb ... used to carry parts small tools samples and f... between floors in buildings with practically ... number of stories Sidewalk elevators travel at ... fpm between the street level and one or two base... ment levels

Oil hydraulic plunger electric elevators are ... signed for low rise light or heavy duty freig... handling They can be installed without spec... building alterations but are limited to buildin... with only a few floors (Fig 3) Use of power... machines imposes severe operating conditions ... elevators Elevator platforms and structures ... subjected to impact loading off balance load ...

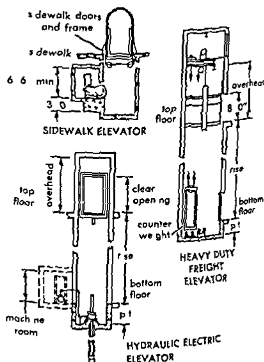


Fig 3 Three types of industrial elevator

and extra static loading To meet these forces that are in addition to load forces, freight elevator design and construction provide greatly increased ruggedness over that of passenger units

Special purpose freight handling elevators are equipped with platforms or arms for carrying specific articles such as rolls of paper, barrels or drums Some of these load and discharge automatically and are arranged so that they operate at any selected floor by means of remote control This is another example of the trend toward more completely automatic operations in connection with materials handling activities *See* MATERIALS HANDLING MACHINES [D O H]

Bibliography American Standards Association *Safety Code for Elevators, Dumbwaiters and Escalators* 4th ed, ASA A17 1 1937, Supplement 1942
Otis Elevator Company, *Vertical Transportation* 1958

Elevator, aircraft

The hinged rear portion of the longitudinal stabilizing surface or tail plane of an aircraft used to obtain longitudinal or pitch control moments The angular setting of the elevator is controlled by the human or automatic pilot through the flight control system (*see* FLIGHT CONTROLS) A typical elevator control surface is shown in Fig 1 Both leading

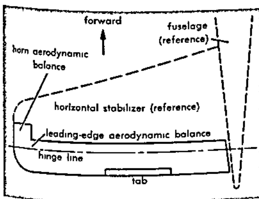


Fig 1 Typical elevator control surface (left hand side)

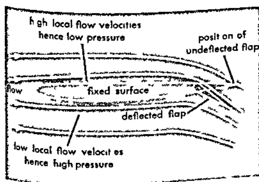


Fig 2 Principle of operation of trailing-edge flap (elevator)

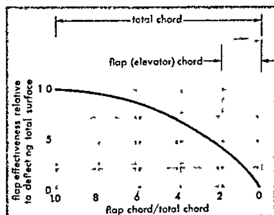


Fig 3 Variation of trailing edge flap (elevator) effectiveness with flap chord to total chord ratio

edge and horn type aerodynamic balances and trailing edge tabs are illustrated These features reduce or eliminate the hinge moments required to deflect the elevator during flight

General principles The operating principles of elevator control surfaces are typical of all trailing edge hinged control devices Deflection of the elevator changes the camber of the entire surface With the trailing edge down high local flow velocities are obtained on the upper airfoil surface and low relative flow velocities are produced on the

The variation of elevator surface control effectiveness with percentage of total chord is shown in

faces

Flutter prevention. Elevator flutter is a divergent oscillation involving one or more degrees of freedom such as rotation about the hinge line and flapping of the main surface (*see* AEROELASTICITY) Hydraulic dampers mass balancing and structural stiffness are employed to prevent elevator flutter

Flight maneuvers The elevator is used to perform pitching maneuvers or maneuvers in which the aircraft's plane of symmetry is not disturbed These maneuvers include air speed adjustments and acceleration normal to the flight path (pull ups or push downs) The elevator also serves to adjust the aircraft's attitude with respect to the ground for take off and landing

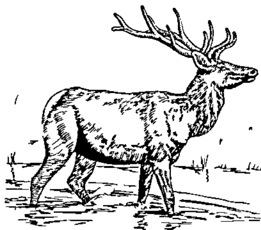
Special applications All moving horizontal stabilizers replace elevator control surfaces on supersonic aircraft and missiles to avoid large losses in effectiveness (*see* STABILIZER) The elevator may be geared to move at a fixed ratio to the deflections of an all moving stabilizer On aircraft without horizontal stabilizers (tailless aircraft), the elevator and aileron surfaces may be combined in surfaces

that operate together as elevators and differentially as ailerons for example elevons [M J AB]

Bibliography L M Milne Thomson *Theoretical Aerodynamics* 1948

Elk

A large American deer *Cervus canadensis* formerly found in most of the open woodlands areas of the United States and southern Canada but now restricted to a few localities primarily the western United States Elk may reach a height of 5 ft and males weigh as much as 900 lb Females are much smaller The males have very large antlers Elk thrive when given protection but they do not disperse readily Therefore a limited number can produce local overpopulations They graze in the uplands during the summer and in the valleys during

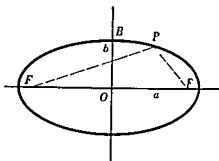


The elk or wapiti *Cervus canadensis* length to 9 ft (From E L Palmer *Fieldbook of Natural History* McGraw Hill 1949)

the winter Starvation may disable or kill many of them in these winter concentrations Elk are excellent food animals and provide good sport for hunters *See* ARTIODACTYLA [J D B]

Ellipse

A member of the class of curves that are intersections of a plane with a cone of revolution (*see* CONIC SECTION) The ellipse is obtained when the plane cuts all the elements of one nappe and does not go through the apex Denote the distance of two points F F of a plane by $2c$ $c > 0$ and let $2a$ be a constant with $a > c$ The ellipse with foci F and F and major axis $2a$ is the locus of points P of the plane such that $PF + PF = 2a$ where PF denotes the distance of P and F This suggests the following construction of an ellipse Put pins at F and F and slip over them a loop of thread of length $2a + 2c$ pulling the thread taut with a pencil If the pencil is moved keeping the thread taut its point traces an ellipse Another way to construct an ellipse is to drill a hole in a stick (at any point other than the midpoint) and move the stick so that its ends slide along two mutually perpendicular lines The point of a pencil inserted in the hole will trace



Ellipse

an ellipse Limiting forms of the ellipse are (1) circle as the two foci approach coincidence (2) the segment (FF) as c approaches a circle be projected orthogonally on a plane parallel to the plane of the circle an ellipse is obtained and every ellipse may be obtained in manner Lines joining the foci to a point P of ellipse make equal angles with the tangent to ellipse at P and consequently light or sound emanates from one focus is reflected to the other focus This property is used in construction whispering galleries'

The midpoint of F F is the center O of the ellipse and the chord through O perpendicular to major axis is the minor axis whose length is noted by $2b$ If B is a point in which the minor a intersects the ellipse then $BF = BF = a$ and $c^2 = a^2 - b^2$ The ratio $c/a = \epsilon < 1$ is the eccentricity of the ellipse (*see* ANALYTIC GEOMETRY) the half chords perpendicular to the major axis multiplied by a/b their extremities will lie on the circle whose diameter is the major axis Hence Δ denotes the area of the ellipse Δ a/b πa that is $\Delta - \pi ab$ The determination of the length L of an ellipse leads to the integral

$$\int_0^{\pi/2} [1 - \epsilon^2 \sin^2 \phi]^{1\ 2}\, d\phi$$

the complete elliptic integral of the second kind It follows that

$$L = 2\pi a \left\{1 - (\tfrac{1}{2})^2 \frac{\epsilon^2}{1} - \left(\frac{1}{2}\ \frac{3}{4}\right)^2 \frac{\epsilon^2}{3} - \left(\frac{1}{2}\ \frac{3}{4}\ \frac{5}{6}\right)^2 \frac{\epsilon^2}{5} - \right.$$

The volume bounded by the surface obtained by revolving the ellipse with major axis $2a$ and minor axis $2b$ about its major axis is $4\pi ab^2/3$ The area of the surface is

$$2\pi a^2 + \frac{\pi b^2}{\epsilon} \ln \frac{1+\epsilon}{1-\epsilon}$$

See ELLIPTIC FUNCTION AND INTEGRAL [L M BL]

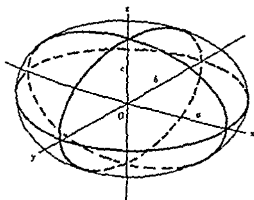
Ellipsoid and spheroid

In analytic geometry the surfaces represented by equations of the second degree in three variables (such as x y z) are called quadric surfaces Ever

quadric surface of finite extent is an ellipsoid Every plane section of an ellipsoid is an ellipse or a single point (provided that a circle is considered

. . . .

whose midpoint is called the center of the ellipsoid All chords of the ellipsoid through the center are called diameters and are bisected by the center If the diameters of an ellipsoid are all equal the ellipsoid is a sphere If the ellipsoid has just one *longest diameter* of length $2a$ and one shortest one of length $2c$ and if the diameter perpendicular to both has length $2b$ these diameters are called principal axes of the ellipsoid, and the volume of the ellipsoid is $4\pi abc/3$ The planes of pairs of axes



Oblote spheroid

are planes of symmetry of the ellipsoid and the equation of the ellipsoid referred to the principal axes as coordinate axes is $(x/a)^2 + (y/b)^2 + (z/c)^2 = 1$

Ellipsoids that have more than one diameter of greatest length are called *oblate spheroids* ($b = a$) and those having more than one diameter of smallest length are called prolate spheroids ($b = c$) An oblate or prolate spheroid is a surface of revolution obtained by revolving an ellipse about its minor or major axis respectively

The earth s surface is approximately an oblate spheroid (actually slightly pear shaped) whose polar diameter of 7900 statute miles is about 27 miles shorter than its equatorial diameters Planes through the polar axis cut the earth s surface in meridians that are more nearly ellipses than circles See ANALYTIC GEOMETRY, EARTH SURFACE AND SOLID OF REVOLUTION [J S F]

Elliptic function and integral

In a certain sense elliptic integrals are the simplest integrals not expressible in terms of elementary functions, elliptic functions arise as the inverse functions of certain elliptic integrals

Let R be a rational function of x and y and set $I = \int R(x\,y)\,dx$ I can be expressed in terms of elementary functions if y is a polynomial of degree 2 or less in x If y is a polynomial of degree 3 or 4 in x I cannot in general be expressed in terms of elementary functions and is called an elliptic integral (If y is a polynomial of degree 5 or higher I is called a hyperelliptic integral and if y satisfies an algebraic equation whose coefficients are polynomials in x I is called an abelian integral) The standard elliptic functions are analogous to trigonometric functions Trigonometric functions may be defined as the inverse functions of certain integrals of the form I they satisfy differential equations are periodic functions and may alternatively be obtained as the simplest periodic functions The standard elliptic functions are the inverse functions of certain elliptic integrals they satisfy differential equations of order 1 and degree 2 are doubly periodic functions, and may alternatively be obtained as the 'simplest doubly periodic functions

Applications In geometry elliptic functions or integrals arise in determining the length of an arc of an ellipse hyperbola or lemniscate the surface of an ellipsoid geodesics on quadrics of revolution parametric representations of plane cubic curves or more generally curves of genus 1 conformal representation problems and other problems In analysis there are applications to differential equations

grals appear in potential theory both through conformal representations and in the potential of an ellipsoid in the theory of elastica the pendulum in rigid body motion in Green's functions in heat conduction and diffusion theory and many other problems

Reduction of elliptic integrals By suitable homographic substitution $x' = (ax + b)/(cx + d)$ $ad - bc \neq 0$ the elliptic integral I can be reduced to an elliptic integral in which the polynomial y^2 appears in normal form The two customary normal forms are Legendre's normal form $y^2 = (1 - x^2)(1 - k^2x^2)$ where k the modulus is a real or complex number $|k| \leq 1$ and $k^2 \neq 1$ and it is usual to set $x = \sin\phi$ and Weierstrass' canonical form $y^2 = 4x^3 - g_2x - g_3$ where g_2 and g_3 the in

normal form respectively as

$$F(\phi\,k) = \int_0^\phi dt/\Delta(t\,k), \qquad E(\phi\,k) = \int_0^\phi \Delta(t\,k)\,dt$$

$$\Pi(\phi\,n\,k) = \int_0^\phi \frac{dt}{(1 + n\sin^2 t)\,\Delta(t\,k)}$$

with $\Delta(t\,k) = (1 - k^2\sin^2 t)^{1/2}$ and in Weierstrass canonical form as

$$\int \frac{dx}{y} \qquad \int \frac{x\,dx}{y} \qquad \int \frac{dx}{(x - c)y}$$

with $y = (4x^3 - g_2x - g_3)^{1/2}$

In Legendre's theory $K = K(k) = F(\pi/2, k)$ and $E = E(k) = E(\pi/2, k)$ are called the complete elliptic integrals of the first and second kinds respectively. $k' = (1 - k^2)^{1/2}$ is the complementary modulus and $K' = K(k') = F(\pi/2, k')$, $E' = E(k') = E(\pi/2, k')$. Complete elliptic integrals as functions of k satisfy linear differential equations of the second order and are hypergeometric functions of k^2. They also satisfy Legendre's relation $KE' + K'E - KK' = \pi/2$ identically in k.

Periods and singularities Elliptic integrals are many valued functions. Any two determinations of I differ by a sum of integral multiples of certain real or complex numbers the periods. E, F and Π are many valued functions of the complex variable $x = \sin \phi$. All three functions have branch points at $x = \pm 1$, $\pm k^{-1}$ and Π has branch points also at $x = \pm i n^{-1}$. The periods of F are $4K$ and $2iK'$, those of E are $4E$ and $2i(K' - E')$. Since the complete elliptic integrals are real when $0 \leqq k \leqq 1$ the first (second) of these periods is called the real (imaginary) period. Although E and F are many valued functions of x, E is a single valued function of F provided that corresponding values of E and F are obtained by integration over the same path and using the same determination of $\Delta(t, k)$.

Inversion of elliptic integrals Jacobian elliptic functions are determined by inversion of the functional relation $u = F(\phi, k)$. It is usual to write

$$\begin{aligned} \phi &= \text{am } u = \text{am } (u, k) \\ \sin \phi &= \text{sn } u = \text{sn } (u, k) \\ \cos \phi &= \text{cn } u = \text{cn } (u, k) \\ \Delta(\phi, k) &= \text{dn } u = \text{dn } (u, k) \end{aligned}$$

With the additional conditions $\text{sn } 0 = 0$, $\text{cn } 0 = \text{dn } 0 = 1$ it turns out that sn u, cn u, dn u are single-valued analytic functions of the complex variable u and that they are doubly periodic and also regular except for simple poles. Nine additional functions are introduced by the notations $1/\text{pn } u = \text{np } u$, $\text{pn } u/\text{qn } u = \text{pq } u$ where p and q stand for any of the letters s, c, d. Similarly Weierstrass $\wp$ *function* is introduced by writing the relation

$$z = \int_{\infty}^{w} (4t^3 - g_2 t - g_3)^{-1/2}\, dt$$

in the form $w = \wp(z) = \wp(z, g_2, g_3)$ and $\wp(z)$ turns out to be single valued doubly periodic and regular except for double poles.

Doubly periodic functions The term p is called a period of a single valued analytic function $f(z)$ regular except for isolated singularities if $f(z + p) = f(z)$. A nonconstant periodic function is either simply periodic when all its periods are integral multiples of a single period or else doubly periodic when its periods are $2m\omega + 2n\omega'$ where m and n are integers, 2ω and $2\omega'$ are primitive periods and $\text{Im } \omega'/\omega > 0$. Two points of the z plane are congruent if they differ by a period. A parallelogram in the z plane is a period parallelogram if every point in the plane is congruent to exactly one point of the parallelogram. If no singularity or zero of $f(z)$ lies on the boundary of the period parallelogram the parallelogram is called a cell. An elliptic function is a doubly periodic function which is regular except for poles. An elliptic function has a finite number of poles in every cell, the sum of the residues at these poles is zero and the sum of the orders of these poles is called the order of the function. Every elliptic function of order 0 is a constant. There is no elliptic function of order 1. An elliptic function of order $r > 1$ assumes in every cell each complex value r times (counting multiplicity). The difference of two elliptic functions with the same periods, same poles and the same principal parts at each pole is a constant. The quotient of two elliptic functions with the same periods, poles and zeros (including multiplicities) is a constant. All elliptic functions with a given pair of primitive periods form an algebraic field and any two such functions are connected by an algebraic relation. Every elliptic function satisfies an algebraic differential equation of the first order. Every elliptic function possesses an algebraic addition theorem, that is, an algebraic relation connecting $f(u)$, $f(v)$ and $f(u + v)$. Conversely any single valued analytic function of z which is regular except for poles and possesses an algebraic addition theorem is either a rational function of z or a rational function of e^{az} for some a or else an elliptic function.

The simplest nontrivial elliptic functions are those of order 2. Choice of a basic function with two simple poles in a cell leads to Jacobi's functions, choice of a function with a double pole to Weierstrass functions.

Jacobian elliptic functions Write $s = \text{sn } u$, $c = \text{cn } u$, $d = \text{dn } u$, $s' = ds/du$ and so on. The periods, zeros and poles are given in Table 1 in which m and n are integers. These functions possess symmetry properties around the points $u = 0$, K, iK' which are set out in Table 2 and on account of which it is sufficient to study the functions in the parallelogram whose vertices are 0, K, $K + iK'$, and iK'. There is a very large number of identities for these functions. Some of the basic ones are

$$s^2 + c^2 = 1, \quad k^2 s^2 + d^2 = 1, \quad d^2 - k^2 c^2 = k'^2$$

$$s' = cd, \quad c' = -sd, \quad d' = -k^2 sc, \quad s'^2 = (1 - s^2)(1 - k^2 s^2) \quad \text{and so on}$$

$$\text{sn } (iu, k) = i \text{ sc } (u, k'), \quad \text{cn } (iu, k) = \text{nc } (u, k')$$

$$\text{dn } (iu, k) = \text{dc } (u, k')$$

The addition theorem for s is

$$\text{sn } (u + v) = \frac{\text{sn } u \text{ cn } v \text{ dn } v + \text{sn } v \text{ cn } u \text{ dn } u}{1 - k^2 \text{ sn}^2 u \text{ sn}^2 v}$$

Table 1 Properties of Jacobi's elliptic functions

Function	Primitive periods	Zeros	Poles
sn u	$4K$, $2iK'$	$2mK + 2niK'$	$2mK + (2n + 1)iK'$
cn u	$4K$, $2K + 2iK'$	$(2m + 1)K + 2niK'$	
dn u	$2K$, $4iK'$	$(2m + 1)K + (2n + 1)iK'$	

Table 2 Symmetries of Jacobi's elliptic functions

v	$-u$	$2K - u$	$2iK' - u$
sn v	$-$sn u	sn u	$-$sn u
cn v	cn u	$-$cn u	$-$cn u
dn v	dn u	dn u	$-$dn u

There are similar addition theorems for c and d These, in combination with the formulas for sn (iu), for example serve to express sn $(u + iv)$ in terms of elliptic functions of u and v; they provide formulas for sn $(2u)$, sn $(u/2)$, and so on By means of these formulas the values of Jacobi's functions at the points $mK/2 + niK'/2$ (m, n integers) and at the points $mK + inK' + u$ can be obtained

The elliptic functions reduce to elementary functions if one or both of the periods become infinite (degenerate elliptic functions, see Table 3)

Table 3 Degenerate elliptic functions

K	K′	k	k'	sn u	cn u	dn u
∞	$\frac{\pi}{2}$	1	0	tanh u	sech u	sech u
$\frac{\pi}{2}$	∞	0	1	sin u	cos u	1
∞	∞			0	1	1

Weierstrass' functions. With $w = 2m\omega + 2n\omega'$ sums and products running over all pairs of integers m, n except $m = n = 0$, there are these functions Weierstrass' sigma function

$$\sigma(z) = z\Pi\left\{\left(1 - \frac{z}{w}\right)\exp\left[\frac{z}{w} + \frac{1}{2}\left(\frac{z}{w}\right)^2\right]\right\}$$

which is an entire function, Weierstrass' zeta function $\zeta(z) = \sigma'(z)/\sigma(z)$, which is a meromorphic function, and Weierstrass' $\wp$ function, $\wp(z) = \zeta(z)$, which is an elliptic function of order 2 with double poles at $z = 0$ and congruent points The invariants are $g_2 = 60\Sigma w^{-4}$ and $g_3 = 140\Sigma w^{-6}$ Legendre's relation becomes $\omega'\zeta(\omega) - \omega\zeta(\omega') = \pi i/2$ The $\wp$ function satisfies the differential equation

$$\wp'^2(z) = 4\wp^3(z) - g_2\wp(z) - g_3$$

and possesses the addition theorem

$$\wp(u + v) = \frac{1}{4}\left[\frac{\wp'(u) - \wp'(v)}{\wp(u) - \wp(v)}\right]^2 - \wp(u) - \wp(v)$$

It is a homogeneous function of degree -2 in z ω ω' Every elliptic function with primitive periods 2ω $2\omega'$ can be expressed in the form $R_1[\wp(z)] + \wp'(z)R_2[\wp(z)]$ where $R_1(w)$ and $R_2(w)$ are rational functions of w, and there are also representations in terms of zeta and sigma functions Degenerate cases lead to elementary functions

Theta functions. The function

$$\theta(v|\tau) = \sum_{r=-\infty}^{\infty} e^{i\pi r^2\tau + i\pi r(2v+1)}$$

with a fixed τ and Im $\tau > 0$ is an even entire function of v It has period 1 it is multiplied by $e^{-i\pi(2v+\tau)}$ when v is increased by τ, and it has simple zeros at the points $v = m + (n + ½)\tau$ (m, n integers) It is usual to consider 4 theta functions

$$\theta_1(v) = -ie^{i\pi(v+\tau/4)}\theta\left(v + \frac{\tau}{2}\right) \qquad \theta_3(v) = \theta(v + ½)$$

$$\theta_2(v) = e^{i\pi(v+\tau/4)}\theta\left(v + \frac{1+\tau}{2}\right) \qquad \theta_4(v) = \theta(v)$$

$\theta(x/2, i\pi t)$ satisfies the partial differential equation $\partial^2y/\partial x^2 = \partial y/\partial t$ and has a simple Laplace transform Elliptic functions and elliptic integrals can be expressed in terms of theta functions $\tau = \omega'/\omega$ in the case of Weierstrass' functions and $\tau = iK'/K$ in the case of Jacobian functions or Legendre's normal form of elliptic integrals

Transformation theory The set of periods of an elliptic function may be described by various pairs of primitive periods The change from one pair of primitive periods to another pair is called a transformation of the elliptic function or integral The quotient of the primitive periods τ, undergoes a homographic substitution, $\tau = (a\tau + b)/(c\tau + d)$, where a, b, c, d are integers and $D = ad - bc$ is positive and is called the degree of the transformation All transformations of degree 1 form the modular group The study of these transformations is of great theoretical interest, has applications to num

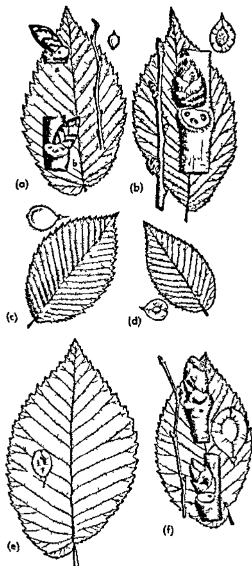
$f(\tau)$ and $f(\tau)$ are algebraically connected when ever τ and τ are connected by a transformation of the modular group See FOURIER SERIES AND INTEGRALS [A ER]

Bibliography P F Byrd and M D Friedman *Handbook of Elliptic Integrals for Engineers and Physicists*, 1954, E T Copson *An Introduction to the Theory of Functions of a Complex Variable*, 1935, A Erdelyi et al *Higher Transcendental Functions* vol 2 1953, A Fletcher, J C P Miller, and L Rosenhead *An Index of Mathematical Tables* 1946, E T Whittaker and G N Watson, *A Course of Modern Analysis*, 4th ed 1940

Elm

Any species of *Ulmus* a genus of hardwood trees in the Northern Hemisphere, with simple serrate, deciduous leaves The American or white elm *U americana*, is the most important species It is a large, typically vase shaped tree which usually grows to about 80 ft but may reach a height of 140 ft The leaves are unequal at the base and doubly serrate, that is the leaf teeth are themselves toothed The winter buds are brown, scaly, and usually at one side of the leaf scar The fruit ripening in late spring is a small, flat winged, elliptical samara fringed with hairs on the edge

It ranges from the eastern half of the United States westward as far as the base of the northern



(a) American elm *Ulmus americana* (b) Slippery elm *U rubra* (c) Rock elm *U thomasi* (d) Siberian elm *U pumila* (e) Scotch or Wych elm *U glabra* (f) English elm *U procera* (A H Graves *Illustrated Guide to Trees and Shrubs* Harper 1956)

Rockies and southward through central Texas to the Gulf of Mexico The tree is also found in southern Canada

The tree is common in moist soil usually at low elevations It is often planted as a shade tree and is perhaps better adapted as a street tree than any other species because the upper leaders of the mature trees rise above utility wires and with their counterparts across the street form a graceful arch

The wood is heavy hard and difficult to split It is used for wagon parts barrel staves shipbuilding furniture flooring sporting goods boxes and baskets The total stand of all species of elm in the United States is given at 8 000 000 000 board ft The annual cut of elm lumber is about 150 000 000 board ft

Slippery elm *U rubra* a smaller tree with larger rough leaves and mucilaginous inner bark is commercially important in the east central United States The uses of the wood are similar to those of American elm

Rock or cork elm *U thomasi* of the northeastern United States is not so common It can be recognized by its usually corky branches small spherical winter buds and flowers and fruit in elongated clusters The wood is similar in nature and use to that of the American elm *See* FOREST AND FORESTRY, TREE [A H G.]

Embioptera

A peculiar order of silk spinning orthopteroid insects related to termites commonly called the embiids or web spinners This order comprises about 1000 species which are chiefly tropical in distribution The body is linear and supple Adults vary in length from 3–25 mm The legs are short with three-segmented tarsi The forelegs are adapted for spinning silk and the hindlegs for reverse locomotion Cerci are short two-segmented and tactile The head is prognathus with orthopteroid mouthparts Metamorphosis is incomplete The females are neotenic and wingless (apterous) Males are usually winged (alate) but in certain genera and species they are apterous The wings are subequal and elongate with the vannal fold obsolete Venation is simple with the veins centered in pigment bands and separated by hyaline stripes The wings are flexible when in repose and folded over the back but are stiffened when extended for flight by the blood pressure in the saclike R_1 (radial) veins Flight is a poorly directed whirling flutter

Plantar surfaces of the mid and basal foretarsal segments bear many hollow silk spinning setae each connected by a duct to a globular syncytial silk gland many such glands are massed in the greatly swollen basal segment Hundreds of silk strands are emitted in unison as the tarsi brush a surface A labyrinth of silk galleries is produced rapidly which constitutes a safe shelter for all embiid activities except adult dispersal All embiid spin silk regardless of the species developmental stage or sex

The galleries radiate on or in the food supply which also constitutes the habitat and consists of bark lichens moss dead leaves or grass Many individuals usually the brood of one female may oc



Body form of a typical embiid (*Parorhagadochir t... chela* Navas female) (E S Ross *Insects Close Up* University of California Press 1953)



Characteristic silk gallery system of an embiid *Haploembia solieri* (Rambur)

cupy one gallery system *See* INSECTA, *see also* ORTHOPTERA [E S R]

Bibliography E S Ross A revision of the Embioptera or web spinners of the New World *Proc US Natl Museum* 94 401–504 1940

Embolism

The sudden blocking of an artery or vein by a clot or other substance which has been brought to its place by the blood current The material carried in the circulation in this process is an embolus Emboli may be composed of thrombi fat air tumor cells masses of bacteria or parasites bone marrow amniotic fluid or atheromatous material from the vessel wall A thrombus which has formed in the heart or one of the vessels is the usual form of embolus Emboli from the right side of the heart or the great venous system of the body come to lodge in the lungs those from the portal system in the liver and those from the pulmonary veins or left side of the heart in some segment of the peripheral arterial tree Pulmonary emboli can result in infarction of the lung However if they are large enough to occlude the main pulmonary artery or one of the major branches the individual may die of shock



Embolus in the pulmonary artery of a dog s lung In this case the embolus was a thrombus (T) which had developed with the formation of new vascular channels (C)

Embolization is a common method of spread of tumor cells which is one reason for the frequency of metastatic tumors in the liver and lungs

Injury to bone or adipose tissue may result in fat embolism Following its escape from the injured fat cell the fat gains access to the venous system It may then come to rest in the capillaries of the lung or it may pass through the lungs to find its way to some other tissue such as the brain or kidney

Air embolism may be a complication of a surgical procedure particularly those about the neck It can also result from rapid decompression as in *caisson disease with the formation of bubbles of* nitrogen in the blood Infected emboli can form new foci of infection at their sites of lodgement

With complete obstruction of a vessel by an embolus an infarct may result *See* INFARCTION [R A V]

Bibliography W A D Anderson (ed) *Pathology* 3d ed 1957 W Boyd *A Text book of Pathology* 6th ed 1953

Embolomeri

A group of Carboniferous and early Permian labyrinthodont amphibians characterized by a vertebral centrum formed of two complete rings—intercentrum and stouter true centrum—both pierced for the notochord The embolomeres are primitive in



Embolomerous amphibian *Eogyrinus* from the Carboniferous estimated length about 15 ft (*After Gregory*)

many features and have sometimes been regarded as the most primitive of labyrinthodonts It now appears more probable that they represent a series of persistently water dwelling fish eating labyrinthodonts related to the reptile ancestors but not themselves directly ancestral to that class Better known members include *Pteroplax* (*Eogyrinus*) of the Pennsylvanian and *Archeria* of the Permian *See* AMPHIBIA FOSSILS LABYRINTHODONTIA [A S R]

Embolus

An abnormal mass or particles carried in the blood stream to a site too small to allow passage so that blood obstruction usually follows A majority of all emboli originate as fragments of thrombi which

have broken off the fixed clot Clusters of tissue cells bacteria or parasites and bubbles of gas air or foreign materials may also become emboli

Venous emboli 95% of which originate in the leg veins may pass through the heart to lodge in the lungs often with catastrophic results These frequently fatal pulmonary emboli originate in patients with poor circulation produced by age lack of activity pregnancy varicosities or following operations and other conditions in which the venous blood is slowed down or immobilized so that an embolus forms

Arterial emboli commonly arise from intracardiac or aortic thrombi In this type lodgement occurs in the brain legs spleen and kidneys and damage may be severe or fatal

Trauma to fatty tissues or the fracture of bones containing fatty marrow may release small fat globules Since most of the droplets can pass singly through various circulatory beds any effects are usually the result of accumulation which occurs most often in the lungs and in the brain with variable results

Air or gas emboli may result from trauma and other causes such as caisson disease in divers A fair amount of air is required to produce fatal results despite popular belief

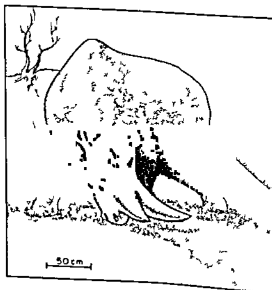
Bacteria may in certain cases form septic emboli which are characteristic of some infections One common method of tumor spread is by means of blood borne malignant emboli which have become dislodged from the parent neoplasm

[E G S T]

Embrithopoda

An order established for the unique mammal *Arsinoitherium* which has been found only in early Oligocene deposits in northern Egypt This animal was of rhinoceros size with a large body and short pillarlike legs There were two huge scimitarlike horns over the nose and two much smaller peglike horns over the eyes As in many herbivorous mammals the cusps of the high crowned teeth were connected to form two transverse lophs on each tooth The exact relationships of this exotic mammal are not clear but it may be related to members of the order Hyracoidea *See* HYRACOIDEA FOSSILS

[W A CL]



Arsinoitherium the early Oligocene embrithopod from Egypt (After C R Knight)

Embryogenesis

The formation of an embryo from the egg showing the basic pattern of organization of the animal with the rudiments of most organ systems at least outlined The axial organization of the embryo (anteroposterior and dorsoventral axes giving the embryo its bilateral symmetry) can often be traced back to the egg before or after fertilization

In addition the presence of colored inclusions sometimes allows detection of the early presence of the precursor cytoplasmic areas of various organs and tissues For example in the egg of the tunicate *Styela* the regions giving rise to skin, gut muscles notochord and nerve cord are roughly outlined just before first cleavage Invisible differences in the fate of different regions of the egg can be detected by experiments such as fragmentation of the egg *See* EMBRYOLOGY EXPERIMENTAL FATE MAPS EMBRYONIC, MEROGONY OOGENESIS *See also* REPRODUCTION PLANT

[G FA]

Embryology

The study of the development of the organism from the zygote or fertilized egg It confines itself mainly to the study of the development of many celled organisms since reproduction in one celled organisms is usually carried out by means other than through the formation of zygotes

The fertilized egg of different organisms differs in size form and organization and is highly characteristic for the species It contains the inclusions typical of all cells and often a nutritive substance called yolk Sometimes pigments and other specialized granular inclusions are localized in particular parts of the cell In some species the cortex differs from the internal portions of the cell and in almost all eggs that have been studied there is a polar distribution of some of the visible inclusions Experimental analysis of development has indicated that even in the absence of visible stratification of inclusions the poles of the egg differ from one another in their developmental capacities and in some cases in their metabolism The importance of the cortex in the regulation of development has also been demonstrated in some forms Polarity and cortical differentiation are only two aspects of the complex organization which characterizes the early developing egg It has been one of the principal tasks of embryology to attempt to describe and analyze this organization those features which distinguish the developing egg of one species from those of another those which distinguish the egg cell from other cells of the organism and especially

those which enable it as a single cell to develop the organized complex multiplicity of the adult

Methods of study Embryology employs many methods to study the organization of the egg and its development In the first instance it attempts to describe the processes of development in terms which are as accurate as possible Descriptive studies have been carried out on eggs and embryos by gross and by microscopic observation In addition phase and electron microscopy have been used with success in some aspects of embryology The description however of embryonic development is a more difficult task than it might seem The study of preserved embryos permits only the static description of successive stages in development and not that of the dynamic processes of change which have produced the differences between one stage and the next Observations on living material allow the investigation of some of the visible aspects of change but many eggs and embryos are so delicate and transparent that [illegible]

of cytochemistry and histochemistry which identify and localize chemical components of particular areas These techniques have been especially fruitful for the study of enzymes Such studies are supplemented by many refined microchemical techniques such as the use of the cartesian diver which allow biochemical analyses of the metabolism of the developing embryo as a whole or of its parts The use of radioactive isotopes has become a useful tool for the biochemical and metabolic study of egg and embryo

Tissue culture It was already clear in the nineteenth century that mere observation of the developing organism was inadequate to permit analysis of the nature and control of developmental organization and experimental techniques began to be devised toward the end of the century for the study of the relationship of the egg or embryo to its environment and of the relations of one part of the developing egg or embryo to another One of the most useful of these techniques has been tissue culture This was devised by R G Harrison who in 1907 performed a crucial experiment which demonstrated that the axon of the nerve cell originates from the developing neuroblast The isolation of embryonic cells in tissue culture indicates the inherent developmental capacities of the cells whether these capacities are greater the same or less than when the cells develop in the whole egg Another useful method has involved the combination of cells from different areas by transplantation See CULTURE TISSUE

Processes of development Embryology classifies the processes of development into a number of overlapping phases

Gametogenesis This is the preparation of the gametes by the parents of the future embryo and in animals deals with the study of the egg and spermatozoon including meiosis and the maturation of their nuclei Gametogenesis has been studied primarily by descriptive cytological histochemical and cytochemical methods In a few cases the origin of the primary germ cells has been studied experimentally See GAMETOGENESIS MEIOSIS

Fertilization This phenomenon involves the union of the two gametes and consequent initiation of development proper Fertilization was described in great detail cytologically at the end of the nineteenth century Knowledge of some of these details is being amplified in the twentieth century through the use of the electron microscope The role of the sperm nucleus as the bearer of hereditary factors contributed by the father was established on cytological grounds at the end of the nineteenth century This aspect of fertilization has also been studied experimentally by hybridization the fertilization of the eggs of one species by the spermatozoa of another In addition nineteenth century experimental studies on fertilization indicated that the spermatozoon plays an additional role beyond the introduction of the paternal chromatin namely the activation of the egg or the initiation of its actual development It was discovered that this role of the spermatozoon could be assumed by many physical and chemical agents differing according to the various eggs used in a process called artificial parthenogenesis The extent to which development can be completed under these conditions depends on the detailed behavior of the nuclear material While artificial parthenogenesis was first demonstrated in eggs fertilized externally the experiment was later performed successfully in the rabbit

The egg has also been fragmented before or after fertilization and its parts then permitted to develop This experiment is known as merogony and its results have indicated that in some forms removal of particular areas of the cytoplasm results in defects of later development while in other species no such defects ensue Androgenetic merogony the fertilization of an enucleated egg fragment of one species by a spermatozoon of another has been useful in assessing the relative roles of cytoplasm and nucleus in the determination of specific developmental and larval characters The reaction which under normal circumstances insures that an egg will be fertilized only by a spermatozoon of its own species is one of great sensitivity and selectivity It has been demonstrated to be similar in character to the interaction between antigens and antibodies and this aspect of the fertilization process has been studied by immunological and serological methods See IMMUNOLOGY SEROLOGY

Related to the fertilization process is a visible redistribution of some of the inclusions of the egg which occurs in many forms This has been described both morphologically and cytochemically and in some cases has been studied experimentally by centrifugation which regroups the inclusions in an atypical manner The centrifugation experiments show that the visible pigments and inclusions can in some cases be relocated without affecti

subsequent development But this method is one of which embryologists are critical since at high speeds centrifugation disturbs development by affecting the invisible ground substance This is the matrix in which the visible inclusions are suspended and results are difficult to evaluate and interpret

Cleavage All developing eggs undergo the process of cleavage or cell division by mitosis in a pattern characteristic for each species The study of cleavage has been primarily descriptive although in some cases experiments separating parts of cleaving eggs from one another has indicated whether the cleavage pattern of partial eggs is similar to that of whole eggs There has been an astonishing paucity of quantitative physical study of the cleaving egg Cleavage has no specific end point since some cells and their division products continue to divide by mitosis throughout the whole adult life of the organism See CLEAVAGE EMBRYONIC

Gastrulation After cleavage has proceeded for a period of hours or days the resultant mass cells often attain a central fluid filled cavity which is then called the blastula Following the blastula the next stage is that of the gastrula Gastrulation is the sum of the processes whereby the cells migrate during the so called morphogenetic movements to attain their final positions in the embryo These movements have been mapped in some forms by the use of vital stains In most multicellular animals, when the movements are completed the cells are arranged in three layers the ectoderm or outer layer the endoderm or inner layer and the mesoderm or middle layer It was believed during the nineteenth century that these germ layers were highly specific in their ability to form tissues and organs but their interchangeability in this respect was subsequently demonstrated by experimental methods to be so extensive that the doctrine of germ layer specificity has been discarded during the twentieth century See BLASTULATION FATE [illegible]

[illegible] ner in their size inclusions pigments and so forth However they are still essentially similar in that they are undifferentiated that is they have not yet attained the structural and functional specialization which marks their adult state The beginning manifestations of their differentiation become visibly apparent usually after gastrulation The process of gastrulation has therefore received great attention from emb[illegible]

[illegible] ied by various methods Isolation of groups of cells in salt solution or in tissue culture and recombination with other cells either in tissue culture or in grafting experiments are examples of such methods

In some forms, for instance some annelids and molluscs the differentiation of cells has been shown to be independent of influences from adjacent cells In others, including vertebrates some cells are susceptible to influences from neighboring cells which induce them to form tissues which they would not form in the absence of those particular neighboring cells The process of embryonic induction has received considerable attention during the first half of the twentieth century and has been studied both experimentally and chemically Attempts have been made to fractionate tissues known to act as inducing agents, and metabolic studies have been made on various groups of cells in both vertebrate and invertebrate embryos in which induction occurs as a means of analyzing the process See EMBRYONIC INDUCTION

Experimental and chemical studies have also been made on differentiation by other methods In some cases biochemical requirements for differentiation have been evaluated by growing embryonic parts in defined media Many modern chemical methods for identifying localizing and tracing substances of metabolic significance have been applied to embryonic stages of organisms Immunological and serological techniques have been used to detect biochemical changes during differentiation In more strictly biological approaches the control of the nucleus over differentiation has been demonstrated by transplantation of the nucleus of one egg into that of another, either of a different age or a different species and by the study of developmental effects of genes in mutations detected by the usual methods of genetics

Morphogenesis Embryologists study not only the differentiation of individual cells but also changes in size and shape of aggregates of cells constituting tissues organs and the whole organism The sum total of these processes is called morphogenesis Morphogenesis or the development of form occurs concomitantly with all the other phases of development which have just been described and includes them all since it deals with differentiation and growth of the whole and its parts Many biologists have been tempted to consider growth and differentiation two different and separable aspects of morphogenesis This is a dangerous oversimplification since in the developing organism these processes are inextricably related components of the processes which develop an organized whole as a result of the influences and properties of organization so uniquely characteristic of egg and embryo See ANIMAL MORPHOGENESIS

Progressive differentiation The constitution and organization of eggs and their specific patterns and procedures of development differ so markedly in different species that embryology has not succeeded in formulating any satisfactory general theory of development The main generalizing concept to have emerged is the principle of progressive differentiation which postulates that the events of each phase of development are intimately related

to events of the phases immediately preceding and following

While the validity of this principle was demonstrated experimentally only in the twentieth century the possibility of such a relationship was suggested by Aristotle The entire history of embryology shows the effort that has been focused on attempts either to confirm or refute that relationship One main argument against the principle was made in the seventeenth and eighteenth centuries when the preformationists claimed that the organs of the adult already existed in a preformed state in the egg or the spermatozoon and then simply unfold as development progresses This abolished the necessity for the concept of development in the usual sense of the word The doctrine of preformation was in turn nullified at the end of the eighteenth century by the concept of epigenesis This occurred when Caspar Friedrich Wolff discovered that the blastoderm of the chick is homogeneous in early stages and attains its later heterogeneous condition by the formation and folding of layers The foundations of later embryology were firmly laid by Karl Ernst von Baer who in the beginning of the nineteenth century extended and elaborated the concept of epigenesis He pointed out that the germ layers first described for the chick by Christian Pander in 1817 are comparable in different forms He emphasized the similarities of embryos to one another and showed that these resemblances are greater the younger the embryos being compared He explained this fact on the basis that development always proceeds from the homogeneous to the heterogeneous, from the general to the special from the type to the individual

Another blow was struck at the concept of epigenesis later in the nineteenth century when Ernst Heinrich Haeckel following the researches of Charles Darwin revived an old concept already refuted by von Baer of parallelism between embryonic stages of higher animals with adult stages of lower ones Von Baer had already shown that embryos resemble only embryos not adults of other species Haeckel however wished to establish causal relationships between the development of the individual and that of the species and expressed these epigrammatically in the catch phrases 'ontogeny recapitulates phylogeny and "phylogeny is the cause of ontogeny His claims that the embryo of a species resembles the adult of its ancestors were false but his generalizations enjoyed great popularity at the end of the nineteenth century The advent of modern genetics made it clear that such resemblances as do exist between embryos can be explained on the basis of the inheritance of common genes In present day experimentation interest in recapitulation has waned in favor of the analysis of specific epigenetic mechanisms without reference to effects produced by the forebears of the embryos studied [J M O]

Bibliography J Needham *A History of Embryology* 1934 J Needham *Biochemistry and Morphogenesis* 1942 A K Parpart (ed) *The Chemistry and Physiology of Growth* 1949 H Spemann *Embryonic Development and Induction* 1938

Embryology, experimental

This term as commonly used designates the science of development of individual organisms from the analytical or causal mechanical point of view Developmental biology implies much the same field emphasizing perhaps the inclusion of plants and microorganisms What is termed chemical embryology although an important aspect is nevertheless only a part of the experimental field There is a tendency to contrast experimental with descriptive embryology it is obvious however that the aim of physiological or mechanical analysis is as much descriptive as is that of purely morphological study The valid contrast is between experimental or analytical embryology and comparative embryology that is directed toward a phylogenetic or evolutionary explanation In the present discussion all studies aimed at understanding the proximal causal factors involved in the individual life history are considered as belonging to experimental embryology even if the experiment involved is no more complicated than the careful visual observation of a portion of the process or a series of chemical analyses at successive stages leading eventually to an interpretation from the causal standpoint

This article attempts to characterize the stages of the life history from the experimental point of view to outline the techniques that have been used in analysis and the concepts that have shaped the ensuing interpretations and to present some report of the generalizations that can reasonably be made

Historical Embryology as a professional field is a nineteenth century discipline Its first practitioners were necessarily engaged in careful description of the morphological steps by which a fertilized ovum becomes an adult organism The need for interpreting the central biological problem of development had of course preceded the tools for accurate observation All working embryologists were from the first engaged in analyzing the embryos they studied either on their own terms or in terms of some allied branch of biology The closest relations were at first with anatomy histology and cytology The cell theory was a major clarification in embryology as well as in all other fields upon which it touched The most important cross disciplinary stimulation came first from physiology later from cell physiology and in the present century from genetics and biochemistry

Germany in the 1870s and 1880s was the site of the first self conscious statement of the problems involved in a causal or physiological view of embryology (Wilhelm His 1874) and of the first program of what was termed Entwicklungsmechanik by its founder Wilhelm Roux The experimental possibilities attracted investigators in many European and American laboratories Experimental embryology appears to have been somewhat of

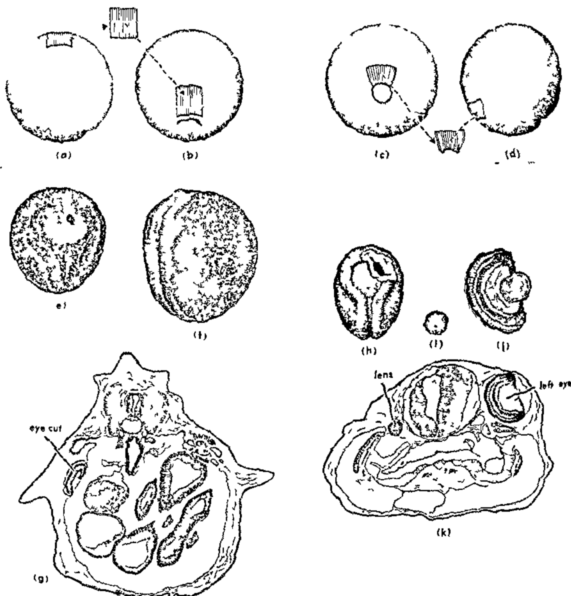


Fig 1 Microsurgery and grafting (a) Presumptive *ectoderm of early gastrula* (b) *Graft* at dorsal lip of blastopore from another germ of the same age (c) *Part of graft* has already been rolled in over the edge of host's *blastopore The part* that is not yet rolled in is now *removed* (d) *Portion* removed is grafted into ventral *side of a third* gastrula (e) Right eye primordium is removed *from the rostral* part of the neural plate of one embryo of *Bombinator* (f) Eye primordium is grafted into *the right* side of another embryo (g) Cross section of the larva developed from the embryo pictured in f an eye-cut has *developed* from the graft (h) Extirpation of the right eye primordium in *Rana esculenta*, at the neural plate stage (i) Lens formed which is smaller than normal (j) Normal eye (k) Cross section of larva 14 days after operation Right eye is absent but a lens *has formed* (From C P Raven, *An Outline of Developmental Physiology* McGraw Hill, 1954)

academic luxury subject flourishing in places where university life has been rich enough to permit free individual research directed by intellectual and aesthetic considerations rather than by economic or ... man centers ... until the att... laboratories ... beca... cent...

group of young embryologists many trained in German universities, grew into the expanding university system in the 1880s and 1890s and founded influential laboratories Thus there has always been a substantial though never disproport... of experimental embryologists ... the academic scene Two ... search centers should be n... Zoologica at Naples and ... Laboratory at Woods ...

marine stations and others patterned after them, investigators found facilities, experimental material in profusion, and scientific companionship

The first, and for more than 50 years the only, journal devoted exclusively to publishing articles on experimental embryology was Wilhelm Roux's *Archiv fur Entwicklungsmechanik des Organismen*, founded in 1895 Embryologists in other countries, as well as many German workers, have in the past preferred to publish their experimental work in less specialized journals, partly in the interest of maintaining contact with other biological fields Thus the original literature is to be found scattered in general scientific journals and in the proceedings of scientific societies as well as in biological serials After World War II embryological journals were started in England Italy, and Japan and by international publishing enterprises Useful reviews are to be found in the usual review journals as well as in the numerous symposium volumes on various aspects of development that now occupy considerable shelf space in biological and medical libraries

Life history The life history of an animal developing from a fertilized egg or zygote may conveniently be divided into three major periods according to the scale and the nature of the events taking place In the first period which includes fertilization cleavage and blastulation the problems are largely cellular in scope In the second period comprising gastrulation and the blocking out of the principal embryonic organs the units are large groups of cells the germ layers or portions thereof which are the rudiments of the future organs Plant morphogenesis in stem or root tips is also on this scale In later animal development as the organs and their coordinating mechanisms come into function in the embryo and fetus the system approximates the juvenile or adult individual and the same physiological or morphological concepts apply *See* BLASTULATION, FERTILIZATION GASTRULATION

Methods The primary technique for the study of developing systems remains observation and the primary tool the microscope Improvement of the microscope in the nineteenth century made possible the inception of modern cell studies subsequent refinements of optical methods have been avidly utilized The following should be mentioned in particular the polarization microscope phase-contrast optics, and photomicrography especially time-lapse cinematography Since World War II the electron microscope has opened a new level of observation *See* MICROSCOPE

Cytological methods of preparing cells and embryos for study have accompanied the use of the microscope The classical fixing agents and dyes were chosen primarily for their freedom from distorting effect on delicate living objects These have been brilliantly supplemented although not replaced by more selective techniques including those that can truly be called cytochemical A highly important method has been that of vital staining Certain dyes not damaging to living protoplasm have been found to persist long enough if applied in localized spots to permit tracing accurately the movements of cells or cell groups *See* FATE MAPS, EMBRYONIC MICROTECHNIQUE

Measurements of various kinds have been essential such as growth in dimensions cell number, or weight over periods of time, growth or increase of parts or components of the organs of embryos of specific chemical content or of enzymatic activity Methods of measuring gas exchange making chemical analyses and physical tests on embryos in early stages of development have exercised the ingenuity of experts in the chemistry and physics of minute systems

In addition to observing and measuring changes it is possible to manipulate the developing system itself such as the germ cells or the zygote with its nucleus cytoplasm and specific milieu One or more components can be eliminated and the developmental effect studied or substitutions or additions can be made in order to test hypotheses Thus by selecting the parent organisms from genetically known stocks it is possible to observe the effect of the presence absence or rearrangement of a single chromosome chromosome segment or gene Conversely the normal milieu (whether sea water, fresh water or special organic milieu) can be analyzed and its components may be removed substituted or increased one by one to assess their

tomeres can be removed isolated or transplanted, or the entire cleaving ovum can be dissociated into a heap of separated cells by exposure to milieus which alter the intercellular cement *See* CENTRIFUGATION (BIOLOGY)

As gastrulation and embryo formation proceed microsurgery becomes more and more the instru

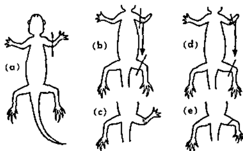


Fig 2 Transplantation of a regeneration bud (a) Amputation of a fore limb If the regeneration bud alone is transplanted onto the stump of an amputated hind limb (d) it will form a hind limb (e), but if part of the original stump (hatched) is transplanted as well (b) a fore limb will be formed by the graft (c) (From C P Raven, An Outline of Developmental Physiology, McGraw Hill, 1954)

ment of choice for analysis The defect experiment tests the effect of removal of one part or rudiment (Fig 1) Numerous types of transplantation can test the rearrangement of parts Among these types of transplantation are autoplastic to the same individual homoplastic from one individual to another of the same species heteroplastic between individuals of the same genus but different species and xenoplastic between more distantly related individuals Grafts or transplants may be made between individuals of different age The site of transplantation may be the identical region from which the transplant was originally removed a nearby location or far removed (Fig 2) The extraembryonic membranes of amniote embryos have furnished useful sites for the growing of grafts The technique of explantation or tissue culture was originally invented for and has been extensively used in this type of study in which an embryonic rudiment is explanted to natural or synthetic culture media Methods have been devised that will permit such an explant to continue development as a whole (organ culture) or to spread out and continue growth as cell or tissue sheets or even as discrete cells (tissue or cell culture)

All of these methods and scaled down methods adapted from work with adult organisms can be applied to embryos in late developmental stages

Fertilization In the micro drama that initiates development the components may be briefly characterized as follows The sperm cell possesses a nucleus containing a single (haploid) chromosome set Its cytoplasmic components are mainly specialized to furnish the means for its brief motile existence although in many species it can be shown that the sperm contributes in addition to its nucleus a division center and possibly other structures to the new organism Although only one sperm cell is required for fertilization in most species fertilization probably never occurs naturally

cautions are required to secure fertilization of a single egg by a single sperm *See* SPERM CELL

The egg is a large even a giant cell Its nucleus is the exact counterpart to that of the sperm containing typically a haploid chromosome set The egg cytoplasm is a highly ordered system with different types of cytoplasm characteristic of different regions The composition and arrangement of these regions varies widely among eggs of different animal groups Such formed structures as yolk droplets mitochondria Golgi bodies and smaller granules are characteristic of most eggs These lie in a matrix of so called hyaloplasm which appears transparent in the light microscope Elements such as pigment granules and "oil droplets are found in eggs of some species (*see* OVUM) These cytoplasmic components are ordered typically in a polar pattern related to density The upper (animal) pole as the egg floats freely is the region where the egg nucleus gives off the polar bodies during

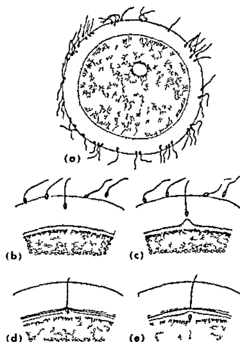


Fig 3 Fert l zat on of the egg of a starf sh (a) The egg n its envelopes surrounded by spermatozoa (b c d e) A sperm p erces the egg envelope and penetrates into the fert l zat on cone The n t al stages of the format on of a per v tell ne cav ty be teen egg and v tell ne membrane *(From C P Raven An Outl ne of Developmental Phys ology McGraw H ll 1954)*

are protected upon release from the parent body by one or more capsular membranous or jelly coverings originating in the maternal ovary or ducts

The morphological and behavioral adaptations of animals for reproduction ensure that at maturity suitable quantities of eggs and sperm are found together in the proper milieu Specific substances are released by each type of germ cell These substances evidently play the role either of dissolving jelly or other substances outside the egg (lysis) or of mutually altering the cell membranes so that the fertilizing sperm adheres by an extruded filament to the egg surface and is eventually engulfed

Both gametes alter subsequent to these events The sperm loses almost all of its meager cytoplasmic structure so that only the nucleus and a very small amount of accompanying material enter the egg cytoplasm The egg undergoes profound changes most of which have been designated as phenomena of activation Starting at the point of contact the egg membrane and the underlying cortex undergo a wave of alteration which travels over the egg surface These changes culminate in the raising of the cell membrane (fertilization membrane) and the establishment of a fluid space between membrane and protoplasmic surface During these changes the egg surface becomes unrecep-

(a) (b) (c)

(d) (e) (f)

Fg 4 (a b c) Sect ons of frog s egg nc ted to develop by pr ck ng w th clean needle show ng the development of the aster near the path of puncture (d e f) Sections of eggs pricked w th needle n presence of blood (double treatment) (From T H Morgan Embryology and Genet cs Columb a Un v Press 1934)

l ve to other sperm The internal protoplasm of the egg partakes in the activation In many animal spe c es the maturation of the egg nu leus is completed only subsequent to sperm entrance Visible rear rangement of cytoplasmic granules and particles may take place In some cases the original po nt of contact serves to set up a visible differential of sym metry so that dorsoventral str icture may be estab l shed at this time The or ginal polarity of the egg is in general retained during this reorganization although in some species the polar axis is subse quently shifted through small angles Various al terations in the metabolism of tl e egg protoplasm and in its protein structure have been detected Sperm and egg nuclei approach one another and may fuse into one or may separate into chromo somes and become arranged directly on the mitotic spindle that has been forming during this period in preparation for the first cleavage d vis on

Artificial parthenogenesis The experiment of artificial parthenogenesis shows that the sperm it self is by no means essential in the activation of the egg Various chem cal alterations in the milieu or phys cal stimulation (Fig 4) can produce cort cal changes in ripe eggs in the absence of sperm Spin dles can then derive from the egg cytoplasm and a certa n percentage of such artific ctivated eggs can complete development with only a mater nal set of chromosomes Thus the mature egg must be thought of as a system containing all the poten tial factors for development and the sperm as a trig ger that sets off the mechanism probably reinforc ing it in various ways and the source of a paternal set of chromosomes

Merogony The experiment of merogony involv ing the fertil zation of egg fragments (Fig 5) shows that development can occur in some species in the absence of cons derable amounts of egg cyto plasm Androgenetic merogony or the fertil zation a lacking the maternal nucleus

he

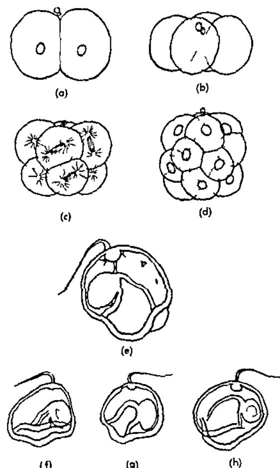
he

es

to

in

t



Fg 5 Cereb atulus (a b c d) No mal cleavage (e) No mal p l d um stage (f g h) P l d a from egg f agments (From T H Morgan Embryology and Ge net cs Columb a Un v Press 1934)

(Fig 6) It was shown statistically that the viable organisms resulting were those possessing at least one chromosome of each pair that would be present in a normal individual Thus the notion of the in dividual ty of the chromosomes and their indispensi ble relation to the future organism was reached on embryological grounds well before the genetic anal ysis of the chromosomes was elaborated It would appear from the many and varied experiments that have been performed on the constitution of the zygote that the irreducible minimum for a develop ing system would be a haploid nucleus within a representative sample of egg cytoplasm including some cortex

Zygote and embryo In the seventeenth and eighteenth centuries ontogeny had been conceived either as the expansion of a miniature organism preexisting in one or the other germ cell (preforma tion) or alternatively as a gradual emergence of new structures from und fferentiated precursors (epigenes s) The d scovery of the chromosomes and their behavior in meiosis and fertilization and the cytolog cal descriptions of the egg cytoplasm that became possible in the 1870s and 1880s per

sperm nucleus forming amphiaster
egg nucleus forming monaster
(a)
male ch omat n el minated
(b)
(c)
male chromat n el minated
(d)
(e)
m tot c sp ndles w thout chromosomes
compact spe m chroma in
haplo d nuclei f om egg nucleus
(f)
(g)

Fig 6 Abnormal fert l zat on (a) Sea urchin egg treated with ether (b) Sea urch n egg fert l zed w h sperm of Myt lus showing fusion of pronuclei (c d) F rst cleavage (e) Four cell stage (f g) Sect on of f og egg fert l zed w th trypaflav n t eated spe m (F om C P Raven An Outl ne of Developmental Phys ology McGraw H ll 1954)

mitted these alternatives to be rephrased as mosaicism and equipotentiality Experimental tests have all tended to show that while neither concept has exclusive validity each represents some aspects of developing systems

Accurate descriptions of morphogenesis had made it clear that the diverse organs of the adult could be traced back to embryonic ones these to regions or areas of the primary germ layers and ultimately to regions or areas of the zygote The doctrine of organ forming germinal areas was enunciated by W His (1874) in his pioneering essays Thus, what was preformed in the germ could be seen to be not a miniature adult but an array of potential areas each of which became cellularized shifted position grew and differentiated into a definite organ according to a pattern and schedule characteristic of the whole system Whether the array of potential areas has mosaic properties is open to experimental testing See ANIMAL MORPHOGENESIS

The first hypothesis offered was based on the discrete nature of the chromosomes (Roux Weismann) The possibility was explored that the zygote nucleus represented a collection of qualitatively different determinants received from the preceding generation and that the mitotic mechanism in cleavage and later operated to sort out determinants in an appropriate pattern to cause each organ forming area to behave in characteristic fashion This was soon shown to be an untenable idea By deformation of cleaving eggs (Fig 7) daughter nuclei can be brought into spatial relations quite different from those produced by the normal lineage of cleavage nuclei Nevertheless such deformed eggs can develop into normal adults Furthermore in certain striking cases of visible nuclear differentiation in early cleavage stages (roundworms, germ cells in insects) the behavior of the nuclear material has been found to depend on a particular region of the egg cytoplasm into which the nucleus wanders at a particular juncture in cleavage Thus if organ forming areas in the early germ are different from one another it is because of the cytoplasm of which they are composed not because different nuclear determinants have been allotted to each one

In the egg cytoplasm granules of various sorts such as mitochondria might be thought of as corresponding to mosaic units This possibility can be tested by cutting off parts of the zygote to see if a partial development follows Another procedure would be the rearrangement of cytoplasmic granules by gravity or centrifugal force Since almost no visible redistribution of cytoplasmic material occurs during cleavage almost but not quite the same purpose is accomplished by waiting until cleavage has partitioned the egg cytoplasm and then separating blastomeres A multitude of such studies on eggs of different species have shown that there are various modes of response to such interventions

In most small marine eggs centrifugation while redistributing visible granules does not produce

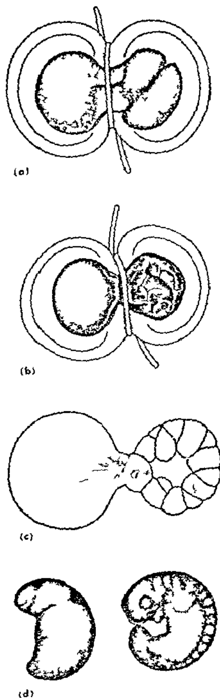
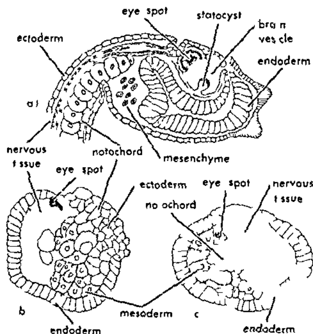


Fig 7 Retarded nucleation in Triton Zygote nucleus pushed to one side by the constriction of the egg (a) First cleavage in the nucleated half (b) Further cleavages in this part the other half is still uncleaved (c) Passage of one nucleus from the nucleated into the nonnucleated half beginning of development in the latter (d) The two embryos that developed from the constricted egg the one produced by the half in which nucleation was retarded is considerably younger but normally built (From C P Raven An Outline of Developmental Physiology McGraw-Hill 1954)



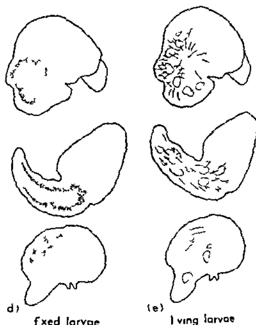


Fig 8 Asc d an larva and *abnormal embryos* (a) An asc d an larva seen from the right (b c) Abnormal embryos formed by centr fuged eggs (d) Fixed larvae (e) Liv ng larvae (From C P Raven An Outl ne of Developmental Phys ology McGraw H ll 1954)

abnormal development *The interpretation* is that the granules so displaced are of no importance to the future organs that the axis of the gastrula and embryo derives from the polar axis of the egg and that this polarity depends on an elastic structure in the ground substance of the cytoplasm that springs back after deformation There are some ex ceptional eggs (Fig 8) in ascidians where cen trifugation displaces visible particles which con tinue to be associated with a definite embryonic tissue Centrifuged embryos therefore are greatly distorted The yolky amphibian egg also can be forced into a greatly distorted pattern of develop ment by centrifugation or simply by holding it with the polar axis inverted and allowing the yolk and other materials to displace themselves by gravity Here the distortion appears to be due to the relation between yolk and protoplasmic masses rather than to the d splacement of one cific type f

the

ing meres after the first cleavage in the amphibian egg where in a certain percent age of cases the fi

age segregates the right from the left side How ever later more exhaustive study showed the situa tion to be much more complicated The presence of the dead blastomere was a decisive factor If the cleaving egg te is divided in half by more delicate means so that each half is free to cont nue develop ment it is possible to produce two perfect em bryos in cases where the first cleavage is indeed median The possibility of producing twins lasts *until* gastrulation Thus the amphibian egg was shown to be not a mosaic but a system capable of regulation *See* EMBRYONIC DIFFERENTIATION

This sort of experiment was repeated with all possible variations on cleaving eggs of many spe cies An isolated blastomere may take one of two opposite courses It may continue to develop as if it were still part of the original egg and form a par tial embryo or it may acquire properties of the whole and form a smaller but complete new in dividual The eggs in which the first result occurs are called mosaic eggs and are believed to have a preformed cytoplasmic pattern before cleavage The eggs of annelids mollusks and roundworms be long to this group Eggs in which a part can form the whole are called regulatory Vertebrates and echinoderms are the principal groups belonging to this category It was the regulatory egg of the sea urchin that gave rise to the notion of the equipo tential harmonic system (H Driesch) when the phenomenon was first investigated

Close investigation of the echinoderm egg by many workers has shown it to be a polarized system and not indefinitely divisible Cytoplasm from ani mal and vegetative (lower) regions must be present in balanced relations if normal development is to ensue This situation has been expressed in the concept of reciprocal gradients of animal and vegetative qualities along the polar axis The po tentiality of forming the micromeres for example or most vegetative cells is not restricted to the cyto plasm which does actually enter into these cells in normal development If this cytoplasm is removed

the ad acent cytopla ...

... ered above the equator of the egg Organ forming areas in general appear to partake at first of this gradient quality the term *field has been used to* indicate a center of developmental potency diminishing *peripherally in all directions* As development progresses the boundaries of each field become restricted *to the actual* ones of the organ in question thus a mosaic situation is achieved The difference between *regulatory* and mosaic eggs is thus reduced to a difference in the time at which this restriction takes place

The *development* of structures from isolated fragments of mosaic systems has given rise to the concept *of self-differentiation* implying that the factors making for differentiation are all located within *the isolate* and to the idea of determination implying that the chain of causal events by which the isolate differentiates is self maintaining These *terms are* of limited utility since they ignore the milieu which is an intrinsic part of the developmental system Thus in the sea urchin egg cited above a certain chemical modification of the sea water in which development occurs results in animalization or the extension of animal qualities into the vegetative hemisphere whereas other *chemicals will* vegetalize the germ suppressing animal derivatives and intensifying the vegetative potencies When transplanted vegetalized animal blastomeres act precisely as if they were vegetative blastomeres Thus determination leads in diametrically opposite directions in different chemical milieus

The role of mitosis in development appears to be in some degree independent of the process of differentiation Cleavage patterns are species and class specific Geometric and spatial rules have been formulated which account convincingly for *these* patterns In the eggs with spiral cleavage and others of the mosaic type cell boundaries delimit the segregated organ forming areas precociously However even in such eggs it is possible to suppress cleavage without entirely suppressing the tendency of the cytoplasmic areas to differentiate specifically In most animals the number of cells comprising an organ rudiment above a certain minimum which evidently is required by the scale of organization of the species protoplasm is subject to much variation Hence the number of mitoses is not fixed One aspect of the species characteristic organization of an embryo is a fairly constant proportion between nuclear volume and cell volume attained as a result of the cleavage process In polyploid animals which unlike polyploid plants are ...

... owth The change in mass is a process usua y commencing only after cleavage is complete Growth may be related or unrelated to cell division In many cases growth is incompatible with differentiation On the other hand some types of differentiation are inextricably bound up with a spec fic cell ...

... of embryos or of parts or components of embryos can be compared with theoretical models Measurement of the increase in chemical components during development is an indispensable prerequisite to understanding metabolism Differential protein growth for example the increase in specific activity of enzymes which characterize adult function is an important measure of tissue and organ differentiation In general the concept of differential or allometric growth relating to body form has illuminated many aspects of morphogenesis of unicellular and multicellular organisms both plant and animal Noncellular structures such as skeletons shells scales or feathers are particularly susceptible to such geometric analysis A highly profitable aspect of the relation of structure to growth has been the analysis of the role of noncellular ground substances in controlling the direction and branching of systems growing in essentially linear fashion such as vertebrate nerve fibers and blood vessels

When the embryo reaches the gastrula stage the more strictly cellular problems of development, while not losing importance become complicated and overlaid by problems in which cell layers cell groups tissues or rudiments are the units of which the interrelations hold the key to the main events Gastrulation in all animals involves radical deformation and rearrangement of the organ forming areas into a layered embryo on which the organ plan of the adult is based Experimental work on these stages largely on vertebrates has involved mapping the changes in shape and movements of the various areas and study of the mechanisms involved in the movements These mechanisms are ...

... and in their relation to the surface coat or superficial intercellular matrix

The new relations of cell groups produced by gastrulation gives opportunity for embryonic induction Some organ forming areas possess the power of autonomous differentiation (determination) before gastrulation Others acquire this power later through contact with adjacent layers Thus the future nervous system (the medullary plate) becomes determined through induction from the mesodermal sheet that grows beneath it The lens of the eye is induced by mesoderm and the eye cup underneath the ear by three adjacent tissues successively Experiments with isolated inductor responding systems and with substitute inductors indicate that the mechanism is at least in part a transfer of chemical material possibly nucleoprotein from underlying to overlying cells *See* EMBRYONIC INDUCTION

Organ forming areas of the embryo after gastrulation have complex properties which cannot be explained as resulting from a single chemical stimulus All such fields appear to have a primary polarity probably related to the original polarity of the egg subsequently supplemented by an axis of symmetry as shown classically in R G Harrison s studies on the amphibian limb As fields become restricted to organ rudiments the details of their future structure become progressively localized and fixed so that they approximate more and more a fine scale mosaic The mosaic situation is seldom completely realized given the capacity for regeneration and repair possessed by all adult organisms In numerous vertebrate organ systems such as the central nervous system gonad kidney and lung there is clearly a cell to-cell induction effect playing a role in cellular differentiation and presumably mediated by transfer of substances In plants various concentrations of the auxins appear to affect cellular differentiation

In arthropods particularly insects secretions of endocrine glands have been shown to control larval growth differentiation and molting (*see* INSECT PHYSIOLOGY) In vertebrates after the establishment of the vascular system the endocrine organs can likewise be shown to come into function as in the adult body and long range transport becomes an important controlling factor

Adult function appears in different organ systems toward the end of the embryonic period Its relation to embryonic differentiation or embryonic function is unclear Nerves and muscles can function before they are fully differentiated structurally Nevertheless an embryo developing under conditions of anesthesia without embryonic reflexes has perfect nervous and muscular systems

One [illegible]

[illegible]

forced and such appears to be the case During every stage of development the germ is a whole organism with homeostatic mechanisms both physiological and *morphogenetic* appropriate to its size and structure (*see* HOMEOSTASIS) A zygote normally starts its career with two of each sort of [illegible]

[illegible] rangement of structures already *present or by* synthesis of new ones Inductions are evidently per [illegible]

[illegible]

hypertrophy after damage are other examples of morphological homeostasis at various stages of development Specific diffusible substances acting at close or long range appear to be at least part of the mechanism

Neoplasia represents the breakdown of morphological homeostasis in late developmental stages A prevailing view of cancer is that it is a cellular phenomenon at its inception and that neoplasia begins in one or a few cells however widespread its effects in the body may later be Research in this field involves not only the study of cellular differences and their possible origin but of the short and long range controls local and systemic that must be overcome before a tumor can grow disproportionately [D R]

Bibliography J S Huxley *Problems of Relative Growth* 1932 J S Huxley and G R De Beer *The Elements of Experimental Embryology* 1934 J Needham *Chemical Embryology*, 1931, J Needham *Biochemistry and Morphogenesis* 1942 H Spemann *Embryonic Development and Induction* 1938 D W Thompson *On Growth and Form* 1941 C H Waddington *Principles of Embryology* 1956 B H Willier P A Weiss and V Hamburger (eds) *Analysis of Development*, 1955 E B Wilson *The Cell in Development and Heredity* 3d ed 1928

Embryonic differentiation

The process by which specialized and diversified structures arise during development of the embryo It is a twofold process involving (1) an increase in the number of cell types and (2) an increase in morphological heterogeneity through the arrangement of cells into increasingly complex structural patterns in the form of tissues and organs *See* HISTOGENESIS

Differentiation begins in most organisms with fertilization of an egg with a sperm after which the relatively large egg divides into many smaller cells called blastomeres The blastomeres receive unequal portions of the cytoplasmic materials of the egg and are therefore initially somewhat different from each other At the end of cleavage the blastomeres are organized into a blastula commonly either a hollow ball of cells or a flattened two layered disk of cells The cells of the blastula lie in different relative positions from those that will be occupied by their descendants in the adult organism (*see* BLASTULATION) By a process known as gastrulation they move to their approximate final positions and are arranged into three basic layers called germ layers However only two layers form in the simpler multicellular organisms (*see* GASTRULATION) The outer layer is the ectoderm from which arise the nervous system and the epidermal layer of the skin The innermost germ layer the entoderm forms the epithelial lining of the digestive tract and contributes the essential tissue of associated organs In all but the most primitive animals a third germ layer the mesoderm is formed by cells which come to lie in the area between the other two layers In higher animals the *mesoderm* gives rise to most of the cells of the organism such as those found in the muscles skeleton blood connective tissue kidneys gonads and certain other organs The molding of groups of embryonic cells into such diverse tissues and organs proceeds

through a variety of morphogenetic processes such as migration aggregation dispersion delamination folding and differential local growth of cells See GERM LAYERS

Cellular differentiation Underlying the visible structural diversification of the embryo is the more fundamental and concomitant process of cellular differentiation (chemodifferentiation) by which embryonic cells are transformed into the highly specialized cells of the adult A characteristic feature of this process is the production of manifold kinds of cells all derived from a single precursor cell The appearance of these new types or strains of cells is not abrupt but is the consequence of a long chain of progressive transformations in the molecular composition and organization of the cell During these transformations each cell gradually loses its capacity to develop in alternative directions but at the same time acquires characteristics essential for further differentiation along its prospective pathway The variety of adult cells produced by the divergent pathways of differentiation varies enormously in different kinds of animals but in each individual the cells cooperatively discharge in an ordered manner the manifold functions of the organism Simpler organisms are constructed of fewer kinds of cells some of which seemingly perform a wider variety of functions than do the cells of more complex organisms However even the most specialized cells carry on a multitude of different chemical activities Depending upon the criteria used hundreds or even thousands of different cell types may be recognized in the more complex animals As our methods of observation and analysis become more refined and penetrating groups of apparently similar cells become resolvable into distinct types in terms either of their functions or of their constituents Special synthetic products commonly distinguish many cell types such as melanin in melanocytes actomyosin in muscle cells and hemoglobin in erythrocytes But even these cell types may be further subdivided Melanocytes of the skin are readily distinguished from those of the retina and muscles of the heart limbs and digestive system constitute distinctive cell strains Moreover the erythrocytes of the embryo have hemoglobin which is distinguishable from that found in adult erythrocytes See HEMATOPOIESIS

Cell structure A cell is an exceedingly complex structure and as such presents numerous possibilities for alteration and diversification Many of these possibilities have been realized in the specialized cells of adult organisms In the center of the cell is the nucleus containing the chromosomes that are responsible for determining the hereditary properties of the cell and for setting the limits of its developmental capacity During cell division the chromosomes are duplicated and distributed in equal sets to the daughter cells (see MITOSIS) thus the cells of an organism are initially equipped with identical sets of chromosomes one set derived from the egg the other from the sperm Whether these chromosomes remain alike as cells proceed through divergent paths of specialization is a moot question Three principal chromosomal constituents are now recognized desoxyribonucleic acid (DNA) ribonucleic acid (RNA) and protein The amount of DNA per chromosome set appears to remain constant in the differentiated cells of an organism but the RNA and protein components vary considerably in amount in different cell types The qualitative characteristics of these chromosomal constituents may also vary but no satisfactory means have yet been devised for investigating them Other constituents of the nucleus the nucleoli nucleoplasm and the enclosing membrane also exhibit differences in various cell strains However the most conspicuous distinguishing characteristics of cells are found in the cytoplasm In the cytoplasmic matrix are imbedded various organelles such as mitochondria ribosomes endoplasmic reticulum centrosome the Golgi complex and a host of enzymes and other chemical substances which by their proportions or type characterize the cell containing them See CELL (BIOLOGICAL)

These components of the cell are organized into dynamic integrated patterns that confer on the cell the capacity to function and multiply Moreover once a terminal state of differentiation in the cell has been reached it is conserved and perpetuated Differentiated cells commonly do not multiply but even when division occurs the differentiated characteristics are preserved throughout descendant cell generations Diverse cell strains are organized into still more complex patterns in the form of various tissues which in turn are built up into organs Tissues and organs vary greatly in the complexity of organization of their constituent cells but are generally composed of populations of heterogeneous cells held in rigid patterns by their mutual affinities antagonisms or both The structure and function of tissues and organs undoubtedly are dependent upon the properties of the component cells and likewise the cells are regulated in their function and course of differentiation by the cell population of which they are a part

Mechanisms of differentiation The mechanisms by which the course of cellular differentiation is realized are not precisely known The factors involved may however be divided into two classes (1) intrinsic those operating within the cell and (2) extrinsic those brought to bear upon the cell from outside Both classes of factors play a role in the differentiation of every cell However the relative importance of these factors varies considerably from one cell strain to another and also within the same cell at different stages in its development The fertilized egg begins development with a rich endowment consisting of a nucleus with a set of paternal and maternal chromosomes together with a complexly organized cytoplasm The activation of the egg by the sperm sets off a chain of actions and reactions that progressively transform the p

and chemical constitution of each descendant cell The emergence of new cell characteristics may be attributed to an oscillating interaction between the intrinsic gene makeup of the cell and the surrounding cytoplasm The dynamic imbalance existing between these interacting components drives the cell along its path of differentiation In certain kinds of invertebrate embryos interactions within each separate cell seem sufficient for guiding differentiation to its terminal state Such embryos exhibit mosaic development By contrast in the embryos of vertebrates and certain invertebrates such as echinoderms influences from adjacent cells are an essential part of the differentiation process These embryos show regulative development Embryos showing mosaic development in which cells differentiate autonomously seem to be the rule among such groups as tunicates molluscs and annelids In these organisms the destruction of a blastomere during cleavage results in a corresponding defect in later stages of development that is the injury is not repaired Moreover isolated blastomeres of such embryos tend to continue development as if they were still a part of an intact embryo Throughout differentiation the cells of these embryos are fixed in their developmental capacity and are able to differentiate in only one direction even when placed in abnormal surroundings This type of differentiation is the product of an unfolding sequence of reactions and segregation of activities that follows a course predetermined in the structure of the egg each differentiating step leads to the next without the intervention of influences from adjacent cells

On the other hand in regulative development the differentiation of cells and tissues is initiated and directed by inductive influences emanating from adjacent cells or tissues (*see* EMBRYONIC INDUCTION) In addition to the guiding influence of one tissue upon another the general physicochemical environment established within a population of cells serves to regulate the further differentiation of the constituent cells These population effects are referred to as embryonic fields or gradients In contrast to the cells of mosaic embryos regulative blastomeres exhibit great developmental plasticity When transplanted to new locations in the embryo the cells respond to their new environment by developing along pathways appropriate to their new location Such developmental plasticity declines as differentiation proceeds and the variety of pathways open to a cell is continuously restricted until the terminal state of differentiation is reached It should be emphasized that all gradations between the extremes of mosaic and regulative development exist and that even the tissues of highly regulative embryos in later stages of development exhibit increasingly autonomous differentiation In all cases however differentiation is a gradual progressive process resulting in the formation of specialized cells from generalized precursor cells Cells when fully differentiated are very stable and only under exceptional conditions of growth in tissue culture or during regeneration of injured tissues do cells lose their differentiated attributes and sometimes acquire the characteristics of other mature cells *See* CELL LINEAGE, EMBRYONIC ORGANIZER

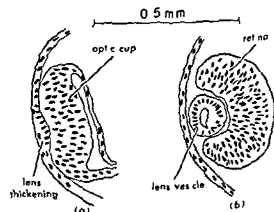
[C L M]

Bibliography E DeRobertis W W Nowinski F A Saez *General Cytology* 2d ed 1954 B H Willier P A Weiss V Hamburger (eds) *Analysis of Development* 1955

Embryonic induction

Each organ or tissue of multicellular animals is formed during embryonic development by a group of cells particular for each organ or tissue In many cases the group of cells destined to form an organ needs an influence from the neighboring cell layer at an earlier developmental stage Such an influence is called embryonic induction In the phenomenon two groups of cells participate the reacting system which provides cell material for the organ to be formed and the acting system whose cells do not take part in formation of the organ but whose presence in the immediate vicinity is necessary for the formation of the organ In most cases the cells of the acting system differentiate into another organ or tissue

One of the established cases of embryonic induction is seen in the lens formation of the salamander s eye In the tail bud stage the rudiment of the lens appears as a thickening of the external ectoderm covering the forebrain of the embryo which is subsequently converted into a lens vesicle This thickening is internally adherent to a layer of the optic vesicle which later develops into the retina If the optic vesicle is removed at an earlier stage lens formation does not occur If the optic cup is grafted in another area of a young embryo that part of the ectoderm which comes in direct contact with the optic vesicle differentiates into the lens Thus the formation of the lens is dependent upon an influence exerted by the optic vesicle or acting system upon the ectoderm the reacting system



Development of the optic apparatus in the salamander (a) An earl er stage (b) Later stage

The reacting system, in each case, has the capacity to react to the influence only during a specific period of development Embryonic induction is operative in the formation of a large number of organs and tissues of the vertebrates For instance, the central nervous system is induced by the archenteron roof, the ear vesicle, by the hindbrain and mesoderm, and the adjacent cartilages by the spinal cord Some cases of embryonic induction are also known in invertebrate animals

Under certain experimental conditions the inductive stimulus can be transmitted without direct contact of the reacting and acting systems Further, specific induction can be caused by some protein and nucleoprotein samples, isolated from differentiated tissues However, the precise mechanism of induction is still a problem to be explored *See* EMBRYOLOGY, EXPERIMENTAL, EMBRYONIC ORGANIZER, NEURAL CREST [T Y]

Bibliography B H Willier, P A Weiss and V Hamburger (eds), *Analysis of Development*, 1955

Embryonic organizer

The area of the vertebrate embryo responsible for induction of the central nervous system and for formation of the complete axial system In the amphibian embryo, this area is the prospective dorsal mesoderm It invaginates through the dorsal lip of the blastopore to become the roof of the archenteron which underlies the dorsal part of the ectoderm (Fig 1) Subsequently, this part of ectoderm develops into the neural plate from which all of the central nervous system including the optic apparatus is derived The dorsal mesoderm itself differentiates into notochord, somites, and head mesoderm Experiments of H Spemann and his coworkers established that the formation of the neural plate in the ectoderm is induced by the dorsal mesoderm (Fig 2) Any intervention in development which disturbs the contact of the ectoderm with the dorsal mesoderm, leads to suppression of neural differentiation

Localized vital staining experiments of W Vogt demonstrated intensive morphogenetic movements occurring in the dorsal mesoderm during gastrula-

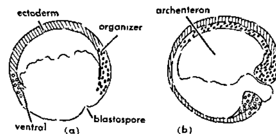


Fig 1 Schemes of gastrulation in the amphibia (a) Median section through an early gastrula before invagination of the organizer area (b) Median section through a late gastrula The dorsal ectoderm lies over the roof of archenteron or organizer

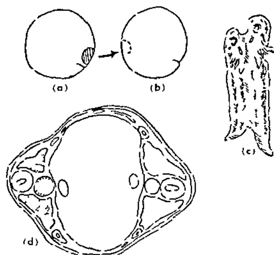


Fig 2 The organizer experiment (a) Donor embryo, an early gastrula in lateral view, with the grafted area indicated (b) Host embryo with the position of the graft indicated (c) Host embryo after culture The primary embryo is on the right and the induced secondary embryo is on the left (d) A cross section through the host embryo, with the primary embryo on the right and the secondary on the left the tissue graft indicated by dotting

tion The neural structures induced by the organizer are often able to induce in turn, sense organs or other tissues They are then referred to as secondary organizers By constriction, transplantation, and explantation a high capacity of regulation was shown for the organizer in its differentiation and induction The area in the embryo of cyclostomes teleosts, birds, and mammals, which corresponds to the amphibian organizer, has also been found to possess both inductive ability and dynamic activity

Despite the large amount of work done in this field, the precise nature of the organizer action is far from clear The question is complicated because the organizer induces not a single tissue, but a large number of different tissues organized into a coherent system Although hypotheses ascribing formation of various tissues either to a large number of inducing factors specific for each tissue or to a gradient of a single factor are supported by facts, the possibility is recognized that combinative effects of a small number of factors cause various types of tissue differentiation All the inductive effects ascribed to the organizer can be simulated by applying various proteins and nucleoproteins isolated from differentiated tissues to the ectoderm However, in the normal organizer, it is still uncertain whether similar substances are at work *See* EMBRYONIC INDUCTION, FATE MAPS, EMBRYONIC, GASTRULATION, NEURAL CREST [T Y]

Bibliography H Spemann, *Embryonic Development and Induction*, 1938, B H Willier, P A Weiss, and V Hamburger (eds), *Analysis of Development*, 1955

Embryophyta

The subkingdom of plants which produce multicellular embryos The zygote (fertilized egg) while retained in the female sex organ divides and forms a mass of cells (the embryo) which is the beginning of a new plant This feature sets apart the Embryophyta from the lower forms of the Thallophyta the other plant subkingdom (*see* THALLOPHYTA)

The Embryophyta consists of two phyla Bryophyta (moss and mosslike plants) and Tracheophyta (vascular plants including club mosses horsetails ferns conifers and flowering plants) *See* BRYOPHYTA TRACHEOPHYTA *see also* PLANT KINGDOM [P D S]

Bibliography H C Bold *Morphology of Plants* 1957 C J Chamberlain *Gymnosperms* 1935 E L. Core *Plant Taxonomy* 1955 H J Fuller and O Tippo *College Botany* rev ed 1954 G M Smith *Cryptogamic Botany* vol 2 2d ed 1955

Emerald

The medium to dark green variety of the mineral beryl Light green stones are properly termed green beryl Among stone dealers the dividing line between what is called emerald and green beryl is sharp differences of opinion on this point are few *See* BERYL

In contrast to the pegmatitic source of most of the beryl varieties emerald is found as a constit

in calcite veins Three groups of mines in Colombia are the government operated Muzo mines the Chivor mines and Somondoco mines The world s finest emeralds come from the Muzo area Other sources include the Ural Mountains India South Africa and Rhodesia A few emerald crystals have been found in North Carolina

In its finest quality emerald is one of the most valuable of gem stones only ruby and diamond are likely to bring higher prices The presence of numerous inclusions fractures or incipient fractures renders most emeralds appreciably more fragile than other varieties of beryl The green color is attributed to the presence of traces of chromic oxide The hardness of emerald is about 7½ on Mohs scale its specific gravity is 2 67 2 85 and its refractive indices vary from 1 565–1 571 to 1 585–1 593 It crystallizes in the hexagonal system *See* GEM *see also* MINERALOGY [R T L.]

Emery

A natural mixture of corundum with magnetite or with hematite and spinel Emery has been used for centuries as an abrasive or polishing material Because the mixture is very intimate and appears to be quite homogeneous it was considered to be a single mineral species until the middle of the nineteenth century The aggregate has a gray to-black color and is extremely tough and difficult to break The specific gravity varies from 3 7 to 4 3 depending upon the relative amounts of the constituent minerals The hardness is about 8 (Mohs scale) less than that of pure corundum which is 9 and is more dependent upon the physical state of aggregation than on the percentage of corundum *See* CORUNDUM

Since early times emery has been recovered from Cape Emeri on the island of Naxos and from other islands in the Grecian archipelago Here it occurs as irregular beds and lenses and in loose blocks associated with crystalline limestone and schists It is also found at several localities in Asia

nineteenth century at Chester Massachusetts where it was associated with diaspore margarite and chloritoid Because of its magnetic properties resulting from the admixed magnetite the Chester material was first worked unsuccessfully as an iron ore Only after the similarity of the associated minerals with those of the Naxos emery was noted was its true nature determined

Although synthetic abrasives have replaced emery in many of its earlier uses it is still used as an abrasive and polishing material by lapidaries and in the manufacture of lenses prisms and other optical equipment Emery wheels emery paper and emery cloth are used not only by lapidaries but also by machinists in the grinding and polishing of steel *See* ABRASIVE GEM CUTTING GRINDING MAGNETITE [C S HU]

Emissivity

The ratio of the radiation intensity of a nonblack body to the radiation intensity of a black body This ratio which is usually designated by the Greek letter ϵ is always less than or just equal to one The emissivity characterizes the radiation or absorption quality of nonblack bodies Published values are readily available for most substances Emissivities vary with temperature and also vary throughout the spectrum For an extended discussion of black body radiation and related information *see* HEAT RADIATION

There are several methods by means of which the emissivity can be determined The one most commonly used is the cavity method In this technique a fine hole is provided in a radiating surface and the ratio of the radiation intensity from the surface to the radiation intensity from the hole yields the emissivity directly This method is quite accurate One can also use an optical pyrometer to determine the emissivity from the brightness temperatures of the hole and the surface in conjunction with Wien s law of radiation

The total emissivity when introduced into the Stefan Boltzmann law gives the total radiated energy W in joules per square centimeter of the real heat radiator as $W = \epsilon\sigma T^4$ Here T represents the

absolute temperature and σ the radiation constant has the value $5\ 67 \times 10^{-12}$ joule cm^{-2} °K^{-4} This energy is always smaller than the energy radiated by the black body since ϵ is less than 1 For example the total emissivity for tungsten is 0 32 at 2500°C which means that at the same temperature tungsten radiates approximately one third the energy of a black body

The spectral emissivity ϵ (the subscript λ denotes the wavelength) provides information on the energy distribution Any spectral emissivity value is valid only for a narrow wavelength interval The wavelength at which ϵ_λ has been determined is indicated by a subscript for instance $\epsilon_{0\ 655}$ A spectral emissivity of zero means that the heat radiator emits no radiation at this wavelength Strongly selective radiators such as insulators or ceramics have spectral emissivities close to 1 in some parts of the spectrum and close to zero in other parts Carbon has a high spectral emissivity throughout the visible and infrared spectrum exceeding 0 90 in certain portions thus carbon is a good black body radiator Tantalum is the only metal with a spectral emissivity greater than 0 5 in the visible spectrum All other metals have a lower spectral emissivity Tungsten is a relatively good emitter with a spectral emissivity of 0 43–0 47 within the visible region of the spectrum [H C S P J W]

Emotion

Emotion is manifested by characteristic overt behavior and speech by changes in bodily processes such as blood pressure and hormone product on and by subjective feelings in one s self Familiar names for the prominent emotions are excitement love fear anger and depression Since the bodily processes wh ch are responsible for the overt behavioral expression and the subjective feel ngs of emotion are continuous the various emotions can exist in all degrees of intensity and in various mixtures Thus the named emotions are appropriate only for the more intense and singularly pure emotional states A profound feature of emotion is its complexity Emotion involves nearly the entire organism at many levels of neural and chemical integration Its diffuse ramifications intermingle with many other bodily and mental processes For example emotion may disrupt digestion and d stort thinking and judgment It can also organize thinking and behavior toward a specific goal such as mating and put zest into life Emotional expression is shaped by the indiv dual s particular learning wh ch is influenced by the family community and general cultural modes

The scientific study of emotion has been handicapped by these characteristics of complexity by its continuity in intensity and in type and by its great variability in different individuals Because of its importance in shaping human behavior and in regulating the individual s feeling states emotion has been studied by philosophers artists poets clinicians and scientists Only the scientific approach with brief reference to clinical studies will be considered in this article

Clinical studies The clinical study of emotion utilizes both the patient s subjective report of his feeling states and the clinician s observations of bodily processes that can be seen without instruments Sigmund Freud and other clinicians using the method of free association in which the patient reports his introspections have contributed to our knowledge of emotion They found that because the free expression of such emotions as fear (anxiety) anger (hostility) and lust is socially unacceptable specialized methods of self control called mechanisms of defense are developed The defenses against overt expression or consciousness of the unacceptable emotion include profound inhibition (repression) somatization with bodily expression such as hysterical paralysis or a psychosomatic disorder like hypertension or less pathological defenses such as rationalization substitution or compensation

The significant point is that the type and intensity of the expression of emotion may be hidden or disguised Neither the subject himself nor an untrained observer may be able to interpret properly the degree of intensity or the particular quality or type of emotion From a scientific viewpoint even the trained clinician s attempts at interpretation of emotion are unsatisfactory because there always remains the possibility that both the patient and analyst were mistaken in their joint effort of introspection and observation Even with their inherent limitations the clinical studies have contributed greatly to the theory and practical management of emotion

A somewhat more systematic method of measuring emotion is to question the subject carefully not about his emotions but about the symptoms of emotion These symptoms may include such things as pounding of the heart dryness of the mouth butterflies in the stomach and trembling This approach is severely limited because most people are poor self-observers especially during emotional states when they are preoccupied by the arousing situation A more objective method is to observe and record all the aspects of overt behavior that indicate emot on These include such aspects as facial expression pitch and loudness of the voice gestures vigor and jerkiness of movement and trembling The difficulty is that these observable signs are such a small part of the total process and there is such great individual difference in expression that only a moderately accurate estimate of the intensity of emotional arousal and little certainty of the quality or type of emotion can be made

Theories A completely objective and quantitative approach has been attempted by direct measurement of the physiological processes of emotion This bodily approach was provided a scientific basis when William James and G C Lange proposed the theory that the bodily processes of emotion such as increased heart rate and blood

changes precede the awareness of emotion James summarizes his theory in 1890 as follows

Our natural way of thinking about these coarser emotions [such as grief fear rage love] is that the mental perception of some fact excites the mental affection called the emotion and that this latter state of mind gives rise to the bodily expression My theory on the contrary is that the bodily changes follow directly the perception of the exciting fact and that our feeling of the same changes as they occur IS the emotion every one of the bodily changes whatsoever it be is FELT acutely or obscurely the moment it occurs

W B Cannon severely criticized James and Lange s theory on the basis of experimental work with animals Cannon s theory of emotion supported by the work of H Head and P Bard is called the thalamic theory In brief Cannon proposed that the brain region called the thalamus receives the emotion provoking stimulus and either by itself or with the support of impulses from the cerebral cortex activates the bodily processes of emotion and at the same time transmits to the cortex the pattern just released to the periphera thus enabling awareness of the emotion The chief difference between Cannon s thalamic theory and James peripheral theory is that awareness of emotion comes from the thalamic neural response pattern rather than only from the sensory feedback from the peripheral organs as James thought The modern activation theory as summarized by D B Lindsley is essentially an extension of Cannon s thalamic theory to include the contribution of recently discovered brain mechanisms

Neurological and endocrine involvement One may understand the activation theory by referring to Fig 1 which traces the principal central nervous system structures and probable pathways involved in emotional behavior The sensory information coming both from the exteroceptors (eyes ears nose) and the interoceptors (proprioceptors in all muscles tendons and visceral organs) activates the reticular formation and thalamus on its way to the cerebral cortex The cortex and thalamus together via neural feedback circuits make the interpretation of the sensory situation

Depending on the nature of the interpretation the degree and type of emotional arousal is developed by the continuous interaction of all these structures The cortex either can exert additional activation by stimulating still further the thalamus and reticular formation or it can dampen the arousal by inhibitory action In an emergency situation involving such an emotion as fear the activated thalamus via the autonomic nervous system causes the many organ changes of the emotional reaction described by Cannon These include increased adrenal secretion vasoconstriction of the skin blood vessels dilatation of muscle and visceral blood vessels increased blood sugar and clotting element in the blood dilatation of the pupil increased muscular tonus more rapid respiration and cessation of peristalsis These organ changes are sensed and they further support and increase the arousal of the reticular system thalamus and cortex Thus emotional arousal tends to build up if the stimulating situation continues even though the external stimulation may not have increased

Quantitative response pattern With this understanding of the neurological and endocrine involvement in emotion it has been possible to record simultaneously the changes in the appropriate organ systems to obtain an objective quantitative response pattern which can describe the intensity and quality of the emotional arousal A F Ax conducted a study of this type A description by E R Hilgard follows The experimenter attached various devices to his subjects so that he could record at once seven different physiological

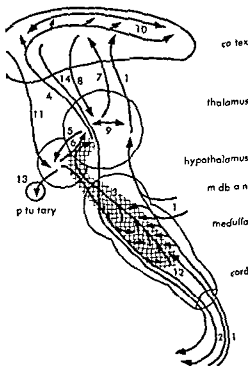


Fig 1 Schematic representation of principal central nervous structures and probable pathways involved in emotional behavior The diagram does not include the cerebellum and certain basal ganglia that may also participate 1 Somatic and cranial afferents 1 direct thalamocortical projections 2 visceral afferent pathways 3 centripetal projections of reticular formation 4 diffuse thalamocortical projections 5 6 interconnections of hypothalamus and thalamus; 7 8 interconnections between thalamus and cortex 9 intrathalamic connections 10 intracortical connections 11 corticohypothalamic pathways; 12 visceral efferent pathways 13 hypothalamohypophyseal tract 14 corticospinal pathways The cross hatched area represents the reticular formation (From S S Stevens Handbook of Experimental Psychology Wiley 1951)

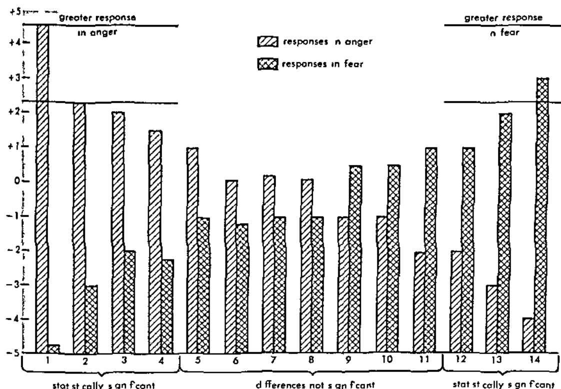


Fig 2 Physiological responses in anger and fear The chart plots changes from the normal (zero) level for 14 indicators all simultaneously recorded The indicators are numbered to correspond to the numbers at the base of the chart 1 Galvanic skin responses (increases in number) 2 heart rate decreases 3 muscle tension increases 4 diastolic blood pressure rises 5 face temperature decreases 6 heart stroke volume decreases 7 heart stroke volume increases 8 hand temperature decreases 9 systolic blood pressure increases 10 face temperature increases 11 heart rate increases 12 muscle tension peaks 13 skin conductance increases 14 respiration rate increases (*From E R Hilgard Introduction to Psychology 2d ed Harcourt Brace 1957*)

indicators of emotional response (pulse rate heart stroke breathing face temperature hand temperature galvanic skin response and muscle action currents just over the eyes) He then on some occasions frightened his subjects and on others angered them In doing this he made very clever use of technicians in his laboratory whose clumsiness invoked fear and whose remarks induced anger The emotional arousal was more natural than is usually the case when a person is strapped up with instruments Each of the 43 subjects was angered once and frightened once (about half in one order half in the other) and then questioned about the reality of the emotion as experienced

The experimenter developed 14 indexes or scores to use in describing the emotional responses of the subjects Half of these in agreement with most findings were common to both fear and anger But the other half seven of the indexes showed significant differences in amount of involvement in anger and in fear (Fig 2) The differences correspond in fear to the action of adrenaline and in anger to noradrenaline as well

These findings on the differences between fear and anger gain support from studies of the adrenal medullas of wild animals D H Funkenstein reported that rabbits which depend for survival on running away as in fear show a predominance of adrenaline and that lions and other aggressive animals whose responses resemble behavior in anger show a relatively high amount of noradrenaline

Such a study appears to demonstrate that at least two intense emotions like fear and anger do have differential physiological response patterns Until similar studies are done involving other emotions it is not certain that all kinds and intensities of emotion are physiologically discriminated at the responding organs It is theoretically possible that the subtle nuances between slightly different feeling states resides only as a brain pattern without peripheral representation With the ever increasing sensitivity and miniaturization of physiological recording transducers detailed knowledge of the emotional states should be greatly increased

[A F A]

Bibliography A F Ax The physiological differentiation between fear and anger in humans *Psychosomat Med* 15 433–42 1953 P Bard On emotional expression after decortication with some remarks on certain theoretical views *Pyschol Rev*

41 309–329 424–449 1934 W B Cannon The James Lange theory of emotions A critical examination and an alternative theory *Am J Psychol* 39 106–124 1927 D H Funkenstein The physiology of fear and anger *Sci American* 1925 74–80 1955 H Head *Studies in Neurology*, vol 2 1920 E R Hilgard *Introduction to Psychology*, 2d ed 1957 W James *Principles of Psychology* 2 vols 1931 S S Stevens (ed) *Handbook of Experimental Psychology*, 1951

Empirical method

The empirical method is not sharply defined It is generally characterized by the collection of a large amount of data before much speculation as to their significance or without much idea of what to expect and is to be contrasted with more theoretical methods in which the collection of empirical data is guided to a large extent by preliminary theoretical exploration of what to expect The empirical method is necessary in entering hitherto completely unexplored fields and becomes increasingly less purely empirical the greater the acquired mastery of the field Successful use of an exclusively empirical method demands a high degree of intuitive ability on the part of the practitioner [P W B]

Emulsion

A dispersion of one liquid in a second immiscible liquid Since the majority of emulsions contain water as one of the phases it is customary to classify emulsions into two types the oil in water (O/W) type consisting of droplets of oil dispersed in water and the water in oil (W/O) type in which the phases are reversed The continuous liquid is referred to as the dispersion medium and the liquid which is in the form of droplets is called the disperse phase

A stable emulsion consisting of two pure liquids cannot be prepared in order to achieve stability a third component an emulsifying agent must be present Generally the introduction of an emulsifying agent will lower the interfacial tension of the two phases

Classification A large number of emulsifying agents are known they can be classified broadly into several groups The largest group is that of the soaps detergents and other compounds whose basic structure is a paraffin chain terminating in a polar group Water soluble soaps (for example sodium or potass[illegible]

[illegible] charge situated at the interface and the counter [illegible]

[illegible] stabilizing action is believed to be similar to that exhibited by certain solid powders For a powder to act as an emulsifier, it must be wetted more by one phase than by the other Whichever phase shows the greater wetting power will become the dispersion medium because such powders congregate at the interfaces and present the greater portions of their surfaces to the liquid which wets them preferentially For example precipitated sulfur which is wetted preferentially by water, stabilizes oil in water emulsions, lamp black, which is wetted preferentially by oil, stabilizes water in oil emulsions Many naturally occurring emulsions, such as milk or rubber latex, are stabilized by proteins Egg yolk proteins stabilize mayonnaise and salad dressing In these and similar types of emulsions stability results from the formation of a protective coating of the protein around each droplet of the disperse phase Certain hydrophilic colloids such as gum arabic or gelatin also stabilize water in oil emulsions by a similar mode of action

Geometrically, the maximum amount of one liquid which can be dispersed in another in the form of spheres of equal size is about 74% of the total available space, independent of the diameter of the spheres However considerably more concentrated emulsions of either type can be obtained because droplets are not necessarily uniform in size and the emulsifying agent permits distortion of the droplets without coalescence The creams used in cosmetics are examples of high concentration emulsions

Properties Various methods are available for determining the type of a particular emulsion but recognition of the following three characteristics is usually sufficient for most purposes (1) The electrical conductivity of an oil in water emulsion is much greater than that of a water in oil emulsion (2) A water soluble dye such as methyl orange will color an oil in water system easily, but will not color a water in oil system For an oil soluble dye such as fuchsin the reverse is true (3) An emulsion will mix perfectly with more of its continuous phase when this is added in pure form

Emulsions may be prepared readily by shaking together the two liquids or by adding one phase drop by drop to the other phase with some form of agitation such as irradiation by ultrasonic waves of high intensity In industry emulsification is accomplished by means of emulsifying machines In a typical machine a mixture of the two liquids containing an emulsifying agent is forced through a narrow slit between a rapidly rotating rotor and a stator The preparation of stable emulsions must be controlled carefully since emulsions are sensitive to such variations as the mode of agitation the nature and amount of the emulsifying agent and temperature changes

The breaking of emulsions is necessary in many industrial operations as for example in the separation of water in oil emulsions in the petroleum industry and in product recovery from emulsions

produced by the steam distillation of organic liq uids Emulsions may be broken by (1) addition of multivalent ions of charge opposite to the emul sion droplet (2) chemical action (addition of acids to emulsions stabilized by soaps) (3) freezing (4) heating (5) aging (6) centrifuging (7) ap plication of high potential alternating electric fields and (8) treatment with ultrasonic waves of low intensity *See* COLLOID LUBRICANT SOAP AND DETERGENT [C S M W O M]

Enamel, nonvitreous

A type of paint characterized by excellent flow and leveling Originally enamels were glossy and simu lated porcelain enamel However the term has been expanded to include semigloss and occasionally flat finishes which level out free from brush marks and other surface irregularities *See* PAINT

[F SD]

Enantiomorph

One of an isomeric pair of chemical compounds whose molecules are nonsuperimposable mirror im ages One molecular configuration of such dissym metric substances is capable of rotating plane polarized light to the right dextro or (+) form while the mirror image rotates the light equally to the left levo or (−) form Each member of such an enantiomorphic pair (optical isomers) possesses identical chemical and physical properties except for interaction with other dissymmetric systems that is other optically active substances plane polarized or circularly polarized light With the dextro form of another dissymmetric system a given (+) form behaves exactly as the corresponding (−) form does with the levo form *See* OPTICAL AC TIVITY [W R V]

Enargite

A mineral having composition Cu_3AsS_4 In places enargite is a valuable ore of copper It is found in orthorhombic crystals but is more commonly co lumnar bladed or massive The mineral has per fect prismatic cleavage metallic luster and grayish black color The hardness is 3 (Mohs scale) and the specific gravity is 4 44 Enargite is one of the rarer copper ore minerals and is found in vein and replacement deposits associated with pyrite galena sphalerite tetrahedrite and bornite It has been mined in Yugoslavia Peru the Philippines and in the United States at Butte Montana and Bing ham Canyon Utah Probably the largest deposit is at Chuquicamata Chile where enargite with other primary copper minerals has been altered to form the great deposit of copper sulfates *See* COPPER COVELLITE [C S HU]

Encephalitis, equine

A mosquito borne viral infection of lower animals which causes disease in horses and mules Two types are known in the United States Western equine encephalitis (WEE) occurs chiefly west of the Mississippi River and Eastern equine enceph alitis (EEE) in the eastern and southern parts of the country Both viruses have been isolated in Canada as well as in South America

The virus a member of arbor virus group A is believed to be maintained in nature by a cycle in birds and mosquitoes If man happens to be bitten by an infected mosquito an encephalitic disease may be produced involving extensive inflammation and destruction of the central nervous system tis sue Inapparent infections are common In the laboratory the virus grows readily in mice chick embryos and tissue cultures *See* CULTURE EM BRYONATED EGG CULTURE TISSUE

The chief vector of WEE is a culicine mosquito *Culex tarsalis* The vector of EEE is not known virus has been isolated from a variety of mosquitoes However the frequency of EEE virus isolation from arthropods is far lower than that of WEE in the respective endemic areas

Venezuelan equine encephalitis (VEE) is dis tributed in northern South America the Amazon valley Trinidad and Panama It is a disease chiefly of equine animals but it may be transmitted to man in whom it usually produces a mild febrile disease *See* VIRUS [J L M]

Encke's comet

A member of the solar system the return of which in 1822 was predicted by T F Encke who had com puted an elliptical orbit for a faint comet observed earlier Encke further showed that the same object had been under observation in 1805 1795 and 1786 and that it represented a new type of object the short period comet The period of 3 3 years still re mains the shortest known for any comet

At its brightest Encke s comet is ordinarily slightly fainter than the naked eye limit The last perihelion passage was on October 20 1957 The other five orbital elements (*see* CELESTIAL ME CHANICS) and the perihelion distance q are listed as follows q 339 a 2 22 AU e 847 i 12° ω 185° and Ω 335°

Encke s comet is responsible for two meteor showers (*see* METEOR) the extended Taurid shower of October and November and a daytime shower the β Taurids discovered by radar observations

Comet Encke is remarkable in one other respect The orbit is systematically becoming smaller and more nearly circular The change is almost entirely in the decreasing value of aphelion The period has decreased by over 2 days in the past cen tury The acceleration has been explained by F L Whipple as resulting from the jet action of gases evaporating from the comet when it is near the Sun The loss of 0 2% of its mass per revolution is sufficient to cause the effect *See* COMET

[R E MC]

Bibliography F L Whipple A comet model I Acceleration of Comet Encke *Astrophys J* 111 375–94 1950

Endocarditis

An inflammation of the lining of the heart or endocardium which also includes the tissues which form the heart valves From the medical standpoint such inflammation is significant mainly when it involves the surfaces of the valves at the exits of the four cardiac chambers The two main forms of endocarditis are bacterial and nonbacterial In the former there is active infection on the surface of a valve This is usually fatal within a period of a few months unless successfully treated by antibacterial drugs The most common cause of nonbacterial endocarditis is rheumatic fever which is sometimes characterized by a slowly progressive scarring of a heart valve After some years this may cause serious mechanical impairment of cardiac function due either to narrowing (stenosis) or incompetence (regurgitation) at the valve orifice *See* BACTERIOLOGY MEDICAL RHEUMATIC FEVER STREPTOCOCCUS [P B BE]

Endocrine gland

A structure which produces and secretes a hormone directly into the circulatory system Endocrine glands are also known as the ductless glands Most arise by segregation and isolation of cells from embryonic epithelium Certain organs function both as exocrine and endocrine structures

Liver Hepatic cells of the secretory tubules which secrete bile into the bile duct system also elaborate important humoral substances Although an antianemic or erythopoietic substance is released by the liver and passes into the vascular system where it is carried through the blood vessels to the site or sites of use doubt still remains as to whether the liver functions as an endocrine gland *See* LIVER

Pancreas The islands of Langerhans develop from segments of the distal portions of the branching ducts These distal segments apparently become isolated from the duct system and transform into the islet cells *See* PANCREAS

Thyroid The thyroid gland arises as an evagination of the entodermal epithelium from the mid floor of the embryonic pharyngeal area between branchial pouches I and II Epithelium separates from the pharynx increases divides into strands then subdivides into masses of epithelial cells The latter masses form the thyroid follicles which become associated to form the thyroid gland *See* THYROID GLAND

Parathyroid Entodermal epithelium migrates away from the dorsal parts of branchial pouches III and IV and becomes associated with the thyroid gland where it forms discrete solid cellular masses *See* PARATHYROID GLAND

Gonads Interstitial cells located between the tubules of the testis function as an endocrine gland (see illustration) In the ovary the interstitial cells between the egg follicles and also possibly a part of the follicle tissue itself produce internal secretory substances In mammals during the postovulatory period the corpus luteum differentiates in the site of the follicle It is an important endocrine structure Similar conditions exist in many viviparous sharks and snakes *See* OVARY TESTIS

Suprarenal (adrenal) body The adrenal cortex is derived from proliferations of cells which migrate inward and away from the surface of the dorsal peritoneal epithelium on either side near the anterior end of the mesonephric kidney in the general body region of the developing stomach and liver Two rounded cortical masses are produced each lying along the anterior inner region of the mesonephric kidney Later the future medullary portions of the adrenal bodies arise from cells which segregate away from the sympathetic ganglia in the general area These sympatheticochromaffin cells migrate into the cortical masses from along the latter s medial side and push inward into the center of each cortical mass to establish the medulla of the adrenal gland

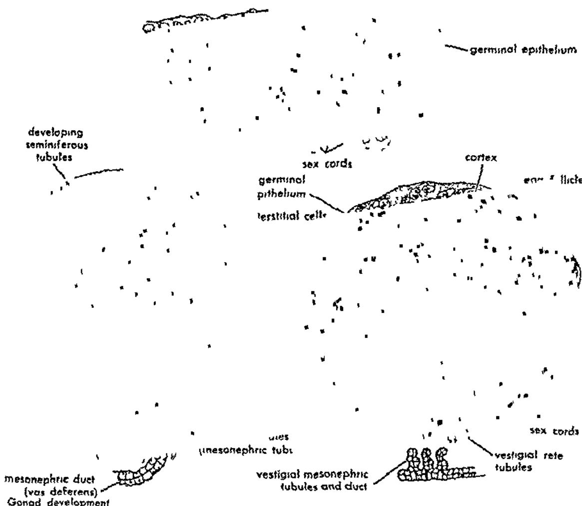
The above description of cortical and medullary portions of the adrenal gland pertains to the mammals In fishes the two components are separate in amphibians they may intermingle or be closely associated and in birds and reptiles they intermingle *See* ADRENAL GLAND

Pituitary gland (hypophysis cerebri) The posterior lobe or pars neuralis is derived from the distal end of the infundibular which pushes down from the floor of the diencephalon The anterior lobe or pars anterior composed of the pars distalis and pars tuberalis together with the intermediate lobe or pars intermedia are derived from the Rathke s pouch The latter is a mid dorsal evagination from the early oral or stomodaeal ectoderm which pushes dorsad and posteriad toward the infundibulum It breaks away from the oral ectodermal epithelium and becomes associated with the infundibular which pushes down from the diencephalic floor *See* PITUITARY GLAND

Kidney The kidney has an endocrine function and secretes the hormone renin The exact site of origin of this secretion is unknown It may be produced by cells associated with the afferent artery of each glomerulus Renin passes into the blood stream and interacts with other substances to produce angiotensin a regulator of blood pressure *See* KIDNEY

Pineal gland This structure develops as an evagination from the roof of the diencephalon of the early embryonic brain It is of dubious function and may produce a substance which suppresses too-early development of the genital organs *See* PINEAL BODY

Paraganglia (chromaffin bodies) These structures are small groups of cells which resemble chromaffin or medullary cells of the suprarenal gland They are associated with various structures such as the abdominal aorta heart kidney and gonads Occasionally these chromaffin bodies are grouped together with the medulla of the adrenal gland as the chromaffin system Their embryonic origin is presumably from cells of the forming



Gonad development

sympathetic ganglia of the autonomic nervous sys tem *See* AORTIC BODY

Thymus. This gland lies in the anterior part of the thoracic cavity in the mammal It is composed of a cortex and medulla The thymus arises from the ventral portion of the third branchial ento-dermal pouch although there may be some con tribution from the fourth pouch also These epithelial contributions from both sides of the em bryonic pharyngeal entoderm migrate posteriad into the anterior thoracic area The thymus gland is found in all vertebrates but its morphology and distribution is variable The thymus is a series of nodules and lies dorsal to the gill slits in fishes whereas in amphibians it consists of two nodules near the angle of the jaws and in reptiles and birds it includes scattered masses in the neck region The function of the thymus is unknown although it is possibly related to sexual maturity *See* THY MUS GLAND [O E N]

Endocrine mechanisms

Those regulatory phenomena in animals or plants which involve as intermediaries one or more hor mones

Hormones These are specific chemical entities secreted by specialized cells tissues or organs, and transported in solution in body fluids to other cells, tissues or organs where they exert a specific physiological action at low concentrations The concept of a hormone was first introduced in 1902 by W Bayliss and E Starling in a study of the coordination of digestive secretion, the term *hormone* was introduced 3 years later by Starling The cells which form hormones are called endo-crine because they pass their secretion into the blood in contrast with exocrine cells which secrete into the digestive tract or other region outside the boundary of the body *See* GLAND

Regulation Regulatory phenomena are those whereby the various activities of the component parts of the organism are modified so that they con tribute to a coherent pattern of activity of the organism as a whole The fact that hormones like vitamins exert their actions at very low concen trations has led to the belief that both kinds of sub stances act as catalysts However, there is as yet no completely established and generally accepted mechanism of action for any hormone and availa ble evidence suggests that unlike vitamins, hor

Properties of hormones (only those hormones which have been reasonably well characterized as distinct physiological and chemical entities are included)

Name	Source or site of formation	Chemical nature	General action or function
Acetylcholine	Peripheral and central nerve endings in many animals	$CH_3COOCH_2N^+(CH_3)_3$	Transmitter substance at certain synapses and nerve endings
Adrenaline (Epinephrine)	Chromaffin cells of vertebrate adrenal medulla and some peripheral sympathetic nerve endings in vertebrates	HO, HO (benzene ring) $CHOHCH_2NHCH_3$	Stimulates glycogen break down to glucose in liver increases rate and force of heart beat erects hairs causes ACTH release
Adrenocorticotrophin (ACTH)	Anterior pituitary of vertebrates	Polypeptide	Stimulates formation and release of steroids from adrenal cortex maintains cortex in normal functional state
Aldosterone	Adrenal cortex of vertebrates	OH CH_2OH; O—CH CO; CH_3; O= (steroid ring structure)	Increases sodium retention by mammalian kidney
Auxin (several substances known indoleacetic acid is typical)	Growing tips of higher plants	N (ring) CH_2COOH	Stimulates increase in size of plant cells formation of roots growth of fruits inhibits growth of buds
Corticosterone	Adrenal cortex of vertebrates	CH_2OH; CH_3 CO; HO; CH_3; O= (steroid ring structure)	Stimulates carbohydrate synthesis and protein breakdown antagonizes insulin
Cortisone	Adrenal cortex of vertebrates	CH_2OH; CO; CH_3 —OH; O=; CH_3; O= (steroid ring structure)	Similar to corticosterone
Ecdyson	Thoracic gland of insects Y organ of crustaceans	Structure unknown	Initiates processes leading to molting

Properties of hormones (Cont.)

Name	Source or site of formation	Chemical nature	General action or function
Estradiol	Follicle cells of vertebrate ovary		Provokes estrus in mammals; proliferation of endometrium in women; stimulates secretion of ICSH by pituitary; mitosis in epidermal cells; inhibits secretion of FSH by pituitary
Estriol	Same		Same
Estrone	Same		Same
Follicle-stimulating hormone (FSH)	Anterior pituitary of vertebrates	Protein	Stimulates growth and secretion of Graafian follicle in ovary; spermatogenesis in testis
Gibberellin (several related substances known)	Many plant tissues		Stimulates elongation of growing plant cells; growth of leaves and fruit
Glucagon	Alpha cells of islet tissue of vertebrate pancreas	Protein	Stimulates breakdown of glycogen to glucose in liver; salt excretion by kidney; inhibits intestinal contractions
Histamine	Injured cells of vertebrates	$NH_2CH_2CH_2C$—NH, CH—N, CH (imidazole ring)	Increases permeability of capillaries; stimulates gastric secretion
Hydrocortisone	Adrenal cortex of vertebrates		Similar to corticosterone
Insulin	Beta cells of islet tissue of vertebrate pancreas	Protein	Increases glucose utilization by cells
Interstitial cell stimulating hormone (ICSH) *luteinizing hormone* (LH)	Anterior pituitary of vertebrates	Glycoprotein	Stimulates hormone production by interstitial cells of gonads of both sexes

Properties of hormones (Cont)

Name	Source or site of formation	Chemical nature	General action or function
Kinetin	Plant tissues	HN—CH₂— O; N, N, CH, N, N, H	Promotes cell division
Noradrenaline	Most peripheral sympathetic nerve endings and some cells in adrenal medulla of vertebrates	HO, HO, $CHOHCH_2NH_2$	Transmitter substance at peripheral postganglionic sympathetic nerve endings constricts arterioles inhibits intestinal contractions
Oxytocin (Lactogogin)	Neurosecretory cells of posterior pituitary vertebrates	Polypeptide	Stimulates milk flow in lactating mammals anti diuretic action in lower vertebrates not mammals large doses stimulate uterine contractions in mammals
Parathormone	Parathyroids of vertebrates	Polypeptide	Stimulates calcification of bones decreases excretion of phosphate
Progesterone	Corpus luteum of mammalian ovary	CH_3, CO, CH_3, CH_3, O=	Stimulates proliferation of uterine endometrium essential for implantation of embryo elicits characteristic behavior of pregnancy
Prolactin	Anterior pituitary of vertebrates	Protein	Essential for milk secretion elicits maternal behavior maintains corpus luteum
Secretin	Walls of vertebrate duodenum	Proteose	Stimulates secretion of pancreatic juice
Serotonin	Many cells and tissues of animals	HO, CH_2CHNH_2, H	Probably a transmitter at certain nerve endings
Somatotrophin (STH) (growth hormone)	Anterior pituitary of vertebrates	Protein	Stimulates tissue growth in young vertebrates increases protein synthesis and fat breakdown inhibits glucose utilization in isolated tissues
Testosterone	Interstitial cells of vertebrate testis	OH, CH_3, CH_3, O=	Stimulates development of male secondary sex characters male behavior and growth of muscular tissue

Properties of hormones (Cont.)

Name	Source or site of formation	Chemical nature	General action or function
Thiamine	Plant leaves	CH_3; N; $\overset{+}{N}$—CH_3; CH_3; N; NH_2; S; CH_2CH_2OH	Stimulates growth of roots and shoots
Thyrotrophin (TSH)	Anterior pituitary of vertebrates	Protein	Stimulates thyroid to form and liberate hormones
Thyroxin	Thyroid of vertebrates	I, I, I, I; HO—⟨ring⟩—O—⟨ring⟩—$CH_2CHCOOH$; NH_2	Increases basal metabolic rate (oxygen consumption heat production) of warm blooded vertebrates stimulates metamorphosis of cold blooded vertebrates molting of amphibians reptiles birds
Triiodothyronine	Same	I, I, I; HO—⟨ring⟩—O—⟨ring⟩—$CH_2CHCOOH$; NH_2	
Vasopressin (antidiuretic hormone ADH)	Neurosecretory cells of posterior pituitary of vertebrates	Cyclic polypeptide	Increases tubular reabsorption of water in mammalian kidney water permeability of amphibian skin Large doses cause arterial constriction

mones are not typically components of enzymes or of enzyme systems Instead they may act by activating or inhibiting enzymes or by modifying the structure or properties of cells or cell components and thus influencing the rate and nature of cellular activities See CATALYSIS VITAMIN

Evidence for hormonal control of a process may take the form of demonstrations that (1) the process is influenced by events occurring elsewhere in the organism when there is no nervous connection involved (2) surgical removal or pathological change in a particular organ or tissue is followed by changes in another organ or tissue elsewhere in the body and (3) extracts of an organ or tissue injected or otherwise administered have effects opposite to those of surgical removal of the organ or tissue or similar to those of pathological hypertrophy of the tissue The final proof of hormonal control after points (1) (2) and (3) have been established comes with isolation purification and chemical identification of the hormone and the demonstration that the pure substance has the same effects as the extract of the tissue of origin In some instances isolation and purification have preceded proof of hormonal function

The types of processes in which hormonal control is involved may be classified conveniently as coordinative conservative progressive and cyclical

Any classification of hormones on the basis of mechanism of action is premature until more is known about such mechanisms

Coordinative control This phenomenon was first demonstrated in the coordination of digestive secretion in the vertebrates When the acidic gastric juice is ejected from the stomach into the duodenum the components of the gastric juice elicit liberation of secretin from the mucosa or inner lining of the duodenum This hormone is carried in the blood to the pancreas where it stimulates the exocrine cells of that gland to form and liberate a secretion poor in enzymes but containing water and salts including sodium bicarbonate, which neutralizes the acid of the gastric juice The duodenal mucosa also forms pancreozymin which stimulates liberation of digestive enzymes into the pancreatic secretion cholecystokinin which stimulates the gall bladder to contract and eject bile into the intestine and enterocrinin which inhibits gastric secretion Contact of food with the stomach causes formation of gastrin which stimulates gastric secretion The effect of these hormonal secretions is to ensure a supply of digestive fluids at the time when food arrives in the appropriate region of the digestive tract and in the case of enterocrinin to terminate gastric secretion when intestinal digestion has begun

Neurohumors Hormonal substances formed by and liberated from special neurosecretory cells in nervous tissue (neurohumors) are often involved in coordinative control In many and perhaps in all cases the transmission of a state of excitation from nerve cell to nerve cell within the nervous system or from nerve cell to muscle cell,

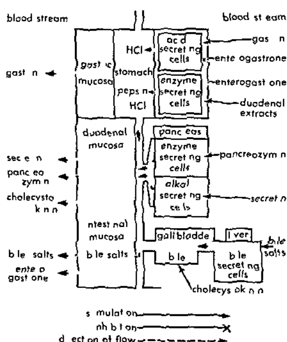


Fig 1 Digestive hormones of the vertebrates (From B T Scheer General Physiology Wiley 1953)

involves a transmitter substance The best known of these is acetylcholine which has been identified as the transmitter between motor nerves and skeletal muscles at parasympathetic nerve endings between nerve cells of sympathetic ganglia in the vertebrates and at some invertebrate nerve endings At the vertebrate motor endplate in skeletal muscle acetylcholine appears to act by increasing the permeability of the endplate membrane to salts as a result there is a flow of ions across the endplate which causes electrical depolarization sufficient to excite the muscle The transmitter at postganglionic sympathetic fibers of vertebrates such as those innervating the arterioles is noradrenaline Nerve endings at which adrenalinelike substances act as transmitters are called adrenergic

Color change Control of color change in animals often involves neurohumors and in crustaceans these substances act as true hormones being liberated in one portion of the body and acting on structures in other parts of the body Animals which can change color in response to visual and other stimuli and thus match their body color or pattern to the background do so by means of *chromatophores* In crustaceans the movements of the pigments within the chromatophores are controlled by neurohumors formed in special groups of neurosecretory cells in the brain and eyestalk Many of these cells have long extensions or axons terminating in the sinus gland a small structure in the eyestalk The secretions are apparently formed in the bodies of the neurosecretory cells and pass along the axon to be liberated from the ending in the sinus gland or elsewhere None of the chromatophorotrophins as these hormones are called has been isolated but partial purification has been accomplished and it is clear that there are several substances each active on one type of chromatophore See CHROMATOPHORE

Color change in the lower vertebrates is under a complex control which varies from species to species In some the only control is by means of intermedin from the intermediate lobe of the pituitary gland This substance has the effect of dispersing the black pigment melanin and thus darkening the animal The hormone is found in all vertebrate pituitaries even though chromatophores are lacking and it may have some function other than its action on melanophores Recent evidence suggests that the effect of intermedin on melanophores may be exerted through an action on the permeability of the melanophore membrane The effects of intermedin are exerted slowly and in some vertebrates color change may be quite rapid In these animals for example the chameleon and some fishes nervous influences are also involved In general cholinergic nerves disperse pigment and adrenergic nerves concentrate it Some vertebrates have only adrenergic nervous control others have both

Conservative control It is well known that hormones exert a conservative control of intermediary metabolism and of salt and water balance in mammals and other vertebrates See METABOLISM

Carbohydrate metabolism Control of carbohydrate metabolism involves the secretion of insulin by the pancreas Removal of the pancreas is followed by a marked increase in the sugar (glucose) content of the blood and by appearance of sugar in the urine The same condition known as diabetes mellitus occurs as a disease in man and may be induced in animals by administering the drug alloxan These effects can be reversed by injection of insulin which is formed in the beta cells of the islets of Langerhans which are small clusters of endocrine cells embedded among the exocrine cells of the pancreas The primary effect of insulin is to increase the utilization of glucose by the cells The mechanism of this action remains uncertain but glucose oxidation and the synthesis of glycogen and fats from glucose are all increased by insulin The pancreas also forms glucagon in the alpha cells of the islets but the physiological significance of this substance is not established See PANCREAS

Experimental diabetes can be alleviated by surgical removal of the pituitary and the injection of extracts of the anterior lobe of the pituitary into normal animals will temporarily increase the blood sugar (diabetogenic effect) The active principles of pituitary extracts act as antagonists to insulin decreasing the utilization of glucose by the tissues The growth hormone somatotrophin STH is the major insulin antagonist in pituitary extracts this hormone also stimulates protein synthesis in the presence of insulin and it is not certain whether or how these actions are related Adrenocorticotrophin (ACTH) also acts as an insulin antagonist in the intact animal but this action is largely indirect and is exerted through stimulation of the cortical

cells of the adrenal glands to form and liberate steroid hormones such as hydrocortisone These steroids stimulate protein breakdown, the synthesis of liver glycogen from protein (gluconeogenesis) and the breakdown of liver glycogen to glucose (glycogenolysis), and hence cause an increase in the blood sugar Removal of or damage to the adrenals is followed by disappearance of glycogen from the liver and inability of the animal to maintain blood sugar levels in fasting or in vigorous activity

In the normal animal, these factors are in balance and the blood sugar is held within well defined limits It is possible that changes in blood sugar level activate the pancreas to increase or decrease insulin secretion It is also possible that the action is mediated through the pituitary STH which antagonizes the action of insulin also stimulates the pancreas to secrete insulin The secretion of ACTH is regulated in part by the level of steroid concentration in the blood, when the steroid level falls ACTH is secreted and this in turn increases the steroid secretion by the adrenals Secretion of ACTH is also influenced by neurosecretory hormones formed in cells in the hypothalamus region of the brain, and transported along axons to the posterior pituitary, where they are liberated into the blood and carried in special vessels to the anterior pituitary

Stress This is a general term applied to a variety of unfavorable conditions to which an animal may be subjected, in many such conditions the secretion of ACTH and of the adrenal steroids is increased, and the resistance of the animal to stress is decreased in the absence of the adrenals Emotional excitement as in fear or anger or vigorous physical activity, also brings into action another emergency control mediated by the medulla of the adrenal glands Nervous stimuli through the sympathetic nervous system cause the medulla to secrete adrenaline, which has its major action in stimulating the breakdown of glycogen in the liver, by activation of the enzyme phosphorylase The result is an increase in blood sugar

Total metabolic level The level of total metabolism in birds and mammals as seen in oxygen consumption and heat production depends on the secretion of the thyroid gland Removal of the thyroid is followed by a decrease in metabolism, injection of thyroid extract or of thyroxine or triiodothyronine increases metabolism Present evidence suggests that these hormones act by increasing the ratio of heat to usable chemical energy produced by cellular oxidations but this requires further confirmation The thyroid is concerned in resistance to cold and thyroid hormones do not usually increase the metabolism of cold blooded animals See THYROID GLAND

The anterior pituitary secretes thyrotrophin (TSH), which stimulates the thyroid to grow and to synthesize and liberate its hormones The level of TSH secretion is in turn determined by the level of thyroid hormone in the blood in a relation similar to that described between ACTH secretion and the level of cortical steroids

Sodium ion concentration The level of sodium ion concentration in the blood and the ratio of sodium to potassium is decreased after removal of the adrenal glands from mammals This change is largely a consequence of an increase in sodium chloride excretion, and a decrease in loss of potassium ion through the kidneys These changes are reversed by aldosterone The secretion of aldosterone is not controlled by the same mechanism which controls secretion of the adrenal steroids active on carbohydrate metabolism It is possible that a separate neurosecretory hormone, formed when the Na/K ratio of the blood decreases, activates secretion of aldosterone The effect of aldosterone may be to decrease cellular permeability to sodium ion and thus increase the effectiveness of the mechanism which pumps sodium ion across cell membranes See KIDNEY DISORDERS

Water balance In mammals, water balance is in part controlled by the posterior pituitary Injection of extracts of the posterior lobe of the pituitary, or of vasopressin antidiuretic hormone ADH causes a distinct decrease in volume of urine formed by a normally hydrated animal Damage to certain neurosecretory cells in the hypothalamus or section of the axons connecting these cells to the posterior pituitary, has the opposite effect inducing the condition diabetes insipidus in which large quantities of dilute urine are produced There is evidence that a neurosecretory product which presumably contains or gives rise to ADH, moves along the axons and is liberated from their endings in the posterior pituitary The ADH probably acts by increasing the permeability of the walls of the kidney tubule to water thus increasing the reabsorption of water by osmosis in the distal portion of the tubule The normal stimulus for ADH release is probably a decrease in the water content (increase in the osmotic concentration) of the blood flowing through the brain See EXCRETION

The calcium and phosphate balance of blood, bone and other tissues is controlled by the parathyroid glands Removal of the parathyroids is followed by a decrease in the urinary excretion of phosphate, increased calcium deposition in bone, and decreased levels of blood calcium Nervous and muscular excitation are increased in consequence of the decreased calcium concentration, and tetany and death follow Injection of parathormone decreases phosphate excretion and causes mobilization of calcium and phosphate from bone, increasing the blood calcium level The normal level of parathormone secretion is probably determined by the level of calcium in the blood, acting directly on the parathyroids

Progressive processes. The progressive processes under hormonal control are those of growth and differentiation in plants and animals

Amphibian metamorphosis The best known instance of hormonal influence on development of animals is seen in amphibian metamorphosis

tadpoles are fed thyroid substance treated with thyroid hormones or even treated with iodine they undergo premature metamorphosis and become miniature frogs If the thyroid or the pituitary is removed from a tadpole metamorphosis fails to occur and the tadpole lacking a thyroid may grow to unusual size without becoming a frog metamorphosis can then be induced by thyroid treatment or by implantation of pituitary glands in a tadpole from which this gland has been removed The thyroid is also concerned in initiating metamorphosis and the downstream migration of anadromous fishes such as the salmon Growth in all the vertebrates depends on the presence of the anterior pituitary and specifically upon somatotrophin The action of this hormone depends in part at least on its ability to promote protein synthesis

swallowing air or more rarely water and the new integument then hardens Before a molt neurosecretory cells in the brain form a secretion which is carried by axons to the corpus cardiacum in the region of the heart and there liberated into the blood This secretion of unknown nature activates the thoracic glands to secrete ecdyson which initiates processes leading to a molt In immature forms larvae and nymphs a juvenile hormone is formed by cells in the corpus allatum near the brain this hormone inhibits the development of adult characters After several larval molts the corpus allatum begins to degenerate in insects with incomplete metamorphosis this degeneration is complete and the last molt leads to formation of the imago or adult In insects with complete metamorphosis the corpus allatum does not degenerate completely and the last larval molt forms a pupa The pupa undergoes extensive reorganization of tissues during which degeneration of the corpus allatum is completed and then molts again to form an imago or adult quite different in appearance from the larva or pupa *See* INSECT PHYSIOLOGY

Crustaceans undergo a somewhat similar although more complicated development but the larval molts have not been studied Unlike insects most crustacean species continue to grow and molt after reaching adult form Ecdyson is formed by the Y organ which resembles the thoracic gland of insects and is essential to molting In many decapod crustaceans neurosecretory cells in the eye stalk form a molt inhibiting hormone which prolongs the period between molts *See* ANIMAL MORPHOGENESIS

Plants The growth and differentiation of plants is also under hormonal control The growing tips of higher plants form auxins which are then transported from the tips toward the base of the plant The auxins have the effect of stimulating elongation of growing cells in proportion to auxin concentration up to a limit higher concentrations may inhibit growth They also inhibit the growth of lateral buds and the abscission of fruits and leaves and stimulate the enlargement of fruit cells and the initiation of new roots The phenomenon of apical dominance whereby the presence of a growing tip inhibits the growth of lateral buds in the basal regions of the stem is attributed primarily to auxin production in the tip and polar transport toward the base The phototropic responses of plants whereby a growing stem tip turns toward the light and a growing root tip turns away and the geotropic responses in which the stem grows upward and the root downward are attributed to lateral polar transport of auxin in stem and root under the influence of light and gravity respectively and the inactivation of auxin by light

The function of two other plant growth hormones remains uncertain Kinetin is formed in many plants and has a stimulating effect on cell division Its effects on intact plants and isolated plant parts depend partly on auxin concentration Gibberellins originally isolated from a fungus which attacks rice plants and other substances of similar nature are also widely distributed Like the auxins the gibberellins stimulate cell elongation and they are particularly effective in causing dwarf varieties of plants to grow to sizes characteristic of normal varieties Unlike the auxins gibberellins move in all directions in the plant and are not transported in polar fashion *See* PLANT MORPHOGENESIS

Sexual differentiation This phenomenon is under hormonal control in a variety of perhaps in all organisms In certain water molds of the genus *Achlya* the sexual development involves at least seven hormones Four hormones two each from the male and female stimulate the male plant to begin forming male organs The male organs in turn secrete a hormone which stimulates the female plant to form female organs The female organs then secrete a hormone which stimulates growth of the male structures toward the female structure and the growing male organ forms a substance which stimulates maturation of the female gametes In the higher plants initiation of flower development depends upon the length of the daylight period Some plants are stimulated to bloom by exposure to at least 12 hours of continuous illumination (long day plants) others bloom only when exposed to at least 9 12 hours of continuous darkness (short day plants) In both types of plants hormones transport the effects of period of illumination to the flower primordia it has been suggested that gibberellin is one of the hormones concerned

Crustaceans have an androgenic gland associated with the testes which is responsible for the differentiation of male secondary sex characteristics and the eyestalks form a hormone which restrains ovarian growth and delays the transformation of males into females characteristic of the life history of certain crustaceans In some crustaceans also the ovary secretes a hormone which induces development of a brood pouch and other temporary female characters In insects the corpus allatum appears to secrete one or more hormones con

Fg 2 Ho monal factors n the human female sexual cycle (From B T Scheer General Phys ology W ley 1953)

cerned with development of female reproductive structures In many other invertebrate animals there are correlat ons of neurosecretory activ ty w th sexual maturation and the discharge of sex ual products In mussels and oysters for example neuro ecretory cells show marked development be fore the maturation of the gametes and their d s charge into the water removal of the gangl a contain ng these cells init ates maturation and spawn ng

Cyclic processes Cyclical processes of repro duct on and molting have been noted in the preced ing section The vertebrate female sexual cycle forms the best known nstance of hormonal control of such a cycle

In mammals the anterior pituitary secretes a gonadotrophin or follicle stimulating hormone FSH which causes development of ovaries in the female testes in the male A second gonadotro phin the luteinizing LH or interstitial cell stimu lat ng hormone ICSH stimulates growth of the interstitial cells of the test s and w th FSH causes maturation of the ovarian follicle and ovulation The interst tial cells of the test

one which induces development of male sexual characteristics and male behavior The ovarian follicle secretes estrogens which induce development of female sexual characteristics As the follicle grows and estrogen secretion increases the estrogen stimulates proliferation of the uterine lining and secretion of more LH by the pituitary The LH in turn causes ripening of the foll cle and ovulation The ruptured follicle becomes a corpus luteum and begins to secrete progesterone which furtl er stimulates growth of the uterine lining If fertilization occurs the fertilized egg is implanted in the uterus and the placenta secretes estrogen progesterone and LH these maintain the corpus luteum and the uterine lining in active condition and inhibit the secretion of FSH by the pituitary If fertilization does not occur the pituitary begins to secrete FSH the corpus luteum breaks down and in the absence of progesterone the uterine lining breaks down completing the cycle

In most mammals the maximum of estrogen secretion is the period of heat or estrus the only time the female is receptive to the male In some mammals such as the rabbit ovulation occurs only following copulation as a result of neurosecretory stimulation to the anterior pituitary leading to LH secretion In other mammals and many other vertebrates as well the female cycle is initiated by an effect of changing lengths of daylight period mediated through neurosecretory cells acting on the [illegible]

anterior pituitary forms prolactin which induces maternal behavior and in mammals is essential to lactation The flow of milk in lactating mammals is initiated by oxytocin *See* ESTRUS LACTATION

[B T S]

Endocrine system

The chemical coordinating system which consists of various specialized glands and cells which elaborate chemical substances called hormones No duct is present to carry these hormones and the endocrine glands are also known as ductless glands in contrast to the exocrine glands Hormones are released directly into the blood stream and serve to integrate the functional activity of cells tissues and organs of the body Certa n glands such as the pancreas function both as endocrine and exocrine glands while others such as the ovary and testis are cytogenic as well as endocrine organs *See* ENDOCRINE GLAND ENDOCRINE MECHANISMS ENDOCRINOLOGY GLAND HORMONE

[C B C]

Endocrinology

The study of the glands of internal secretion and their hormonal products The endocrine glands characteristically secrete d rectly into the blood stream rather than into a duct of any kind The pitu tary or hypophysis thyroid parathyroids adrenals and gonads (either testes or ovaries) are the principal organs of the system along with clusters of cells in the pancreas called the islets of Langerhans *See* ADRENAL GLAND GONAD HORMONE HYPOPHYSIS PARATHYROID GLAND THYROID GLAND

The pituitary is considered the master gland because so many of its hormones are trophic in nature that is they regulate the activity of the other endocrines A complex hormonal balance together with nervous system regulation controls most of the physiologic processes of the body

Some hormones have a specific effect on certain target tissues whereas others have a general effect on most cells of the body In many cases a particular organ may be affected by either synergistic or antagonistic hormone influences from the same or different glands

Chemically the glands produce three basic kinds of hormones the steroids the amines and the protein types but many subtypes and variations both physiologic and abnormal may appear in an individual

an

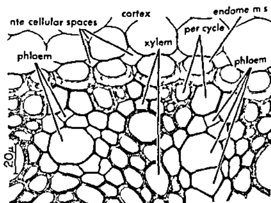
an

actions desirable in therapy [E C ST]

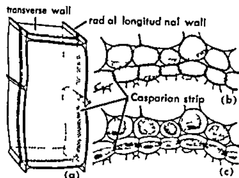
Endodermis

The single layer of plant cells that is located between the cortex and the vascular (xylem and phloem) tissues (Fig 1) It has its most obvious development in roots and subaerial stems In the aerial stem and in leaves it is not always detectable except by histochemical tests However aerial parts of many plants may develop a characteristic endodermis when subjected to below ground growing conditions

The primary phase in the development of the endodermis is identified by a thin band (the Casparian strip) of suberin or ligninlike deposition around each cell in the anticlinal (perpendicular to the surface) walls that is the radial and transverse walls (Fig 2*a*) There are no intercellular spaces in the endodermis and the anticlinal walls



F g 1 Transect on from part of *Zea* root illustrat ng endodermis in relation to cont guous tissues (*K Esau Plant Anatomy W ley 1953*)



Fg 2 Deta ls of endodermal structure (a) Diagram of a cell show ng location of Caspar an strip (b) Cells of endodermis and of ord nary parenchyma before treatment w th alcohol (c) Cells after treatment Casparian str p is seen only in sectional views in (b) and (c) (K Esau *Plant Anatomy* *Wiley, 1953*)

between the cells are blocked by the Casparian strip giving a "watertight appearance to this tissue (Fig 2b) When plasmolized the protoplasts of the endodermis appear to adhere to the Casparian strip The contiguity of protoplast and the material of the Casparian strip (Fig 2c) is supposed by some to prevent the movement of solute ions and water through the walls or between the walls and the cytoplasm and thus restrict this movement to the cytoplasm This relation in turn is thought to play a role in the selective absorption of ions and in maintaining hydrostatic pressure Such a function of the endodermis however has not been proved conclusively

The secondary phase in development of the endodermis consists of a complete suberization of all of the walls This event may be followed by a tertiary phase in which cellulose deposition takes place on the inner tangential and radial walls An oxidation of phenols naphthols and anthrols to quinones occurs in the endodermis of roots and subaerial stems These substances possibly serve as a barrier against the entry of pathogens such as bacteria fungi and nematodes During the tertiary phase of wall development polymerized and oxidized inclusion products are deposited in the cellulose walls See CORTEX PLANT HYPODERMIS PLANT TISSUE SYSTEMS [D S V F]

Bibliography See PLANT ANATOMY

Endopterygota

A division of the insects which comprise the subclass Pterygota The group includes those orders of insects which undergo a complete metamorphosis during their life cycle Therefore they are equally well known as the Holometabola Four stages of development occur during metamorphosis the egg larva pupa and imago or adult The larva commonly called caterpillar grub or maggot hatches from the egg Immature and wormlike it appears in an entirely different form from that of the parent The larva develops into the pupa a quiescent stage in which the insect acquires its adult characteristics The adult emerges from the pupa as a fully developed sexually mature winged insect *See* INSECTA PTERYGOTA [B E R]

Energy

The capacity for doing work Energy is possessed for example by a body that is in motion for stopping it provides work by a compressed or stretched spring for it is capable of doing work in returning to its ordinary configuration by gunpowder or a bomb because of the work it can do in exploding by a charged electrical capacitor for it can do work while being discharged Energy like work is a scalar quantity Its units are the same as those of work and include the foot pound foot poundal erg joule and kilowatt hour *See* WORK

Because a system may possess an enormous store of energy that is not available for doing work energy is better defined as that property of a system which diminishes when the system does work on any other system by an amount equal to the work so done Although energy may be exchanged among various bodies or may undergo transformation from one form to another it has the tremendously important property that it cannot be created or destroyed (*see* CONSERVATION OF ENERGY)

Energy occurs in several well defined forms as kinetic energy potential energy and internal energy these three forms having the common characteristic that they constitute energy stored for possible future use as work and heat which are transient forms of energy that provide the means by which the other forms of energy are transformed or transferred from one body to another Heat is energy in transfer by virtue of a temperature difference existing between the bodies the two methods of transfer without transport of material being thermal conduction and thermal radiation

Internal energy This is the energy present within a body or system such as a fuel steam or compressed gas by virtue of the motions of and forces between the molecules and atoms of the body or system Internal energy is sometimes erroneously referred to as the heat energy of a body It is a property of any given state of a system is evidenced by certain other properties of the system notably temperature and is to be distinguished from any kinetic or potential energy possessed by the system as a whole in its relation to other systems According to the first law of thermodynamics the change of internal energy in any given process is equal to the difference between the heat gained by the system and the work done by the system on other systems external to it *See* INTERNAL ENERGY, THERMODYNAMIC PRINCIPLES

Kinetic energy This is the term applied to the capacity for doing work that matter possesses because it is in motion As everyday experience shows the more massive a body is and the higher its speed the more work it will do upon striking and being slowed down by an obstacle and hence the larger is its kinetic energy Specifically for a body of mass m moving with a speed v the kinetic

E_k is given by

$$E_k = \tfrac{1}{2}mv^2 \qquad (1)$$

Thus if a car of mass 60 slugs (about 1900 lb) is moving with a speed of 90 ft/sec (about 60 mi/hr) its kinetic energy relative to the ground is 243 000 slug ft²/sec² or 243 000 ft lb Now the change in the kinetic energy of a body during a given displacement is equal to the work done by the net or resultant force applied to the body during this displacement a statement that is usually referred to as the work kinetic energy theorem Thus to bring the car in the example to a stop in say 100 ft would require an average retarding force of 2430 lb or about 1 2 tons of force It will be noted that the value of the kinetic energy is always dependent on the body s speed relative to some chosen reference body Thus the kinetic energy of a car relative to a man sitting in it is zero only if the car changes speed relative to the man can it do work on him because of its motion

To derive Eq (1) suppose that a body of mass m and moving with speed v is brought to rest within a distance s by applying to it a constant force of magnitude f The work done is fs and by the work kinetic energy theorem is equal to the change in the body s kinetic energy E_k By Newton s second law of motion $f = ma$ and since the acceleration a is assumed here to be constant $s = v^2/2a$ Therefore $E_k = fs = ma \cdot v^2/2a = \tfrac{1}{2}mv^2$ This expression for the kinetic energy is valid for all speeds except those comparable to the speed of light

The kinetic energy of rotation of a body that is turning about a fixed axis with angular velocity of magnitude ω (radians/sec) is given by $E_k = \tfrac{1}{2}I\omega^2$ where I is the body s moment of inertia with respect to the axis in question

Potential energy This form of energy as contrasted with kinetic energy is the capacity to do work that a body or system has by virtue of its position or configuration Thus elastic potential energy is possessed by a coiled spring that is compressed or stretched (Fig 1) indeed frictional forces in a spring are often so small that more than 99% of the work of deformation is recovered when the spring is released (*see* HOOKE S LAW) Gravitational potential energy is possessed by a body that

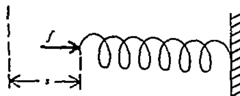


Fig 1 A compressed spring possesses potential energy Here s is the distance that the spring has been compressed from its normal length and f is the force of compression It can be shown from Hooke s law that the elastic potential energy of the compressed spring is $\tfrac{1}{2}ks^2$ where k is a constant known as the stiffness coefficient of the spring

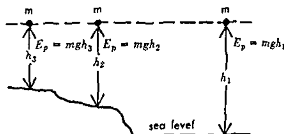


Fig 2 The gravitational potential energy E_p of a body of weight mg for different reference levels

has been raised above the earth s surface for the body can do work in falling to the ground it can strike and drive a nail or a pile or can compress a spring Electrical magnetic chemical and nuclear systems may also possess potential energy

In general the potential energy of one configuration of a system relative to another configuration of it may be defined as equal to the work done against the conservative forces of the system when its parts change from the one configuration to the other Conservative forces are forces such as those of gravity or the force exerted by a spring where the work is recoverable that is the net work done in a round trip is zero (*see* FORCE) As this definition implies the potential energy of a system when in a particular configuration must always be computed with respect to some other arbitrarily selected configuration or position of the system moreover its value is a function only of the initial and final positions and not of the paths followed by the parts in changing position

Gravitational potential energy Suppose that a body of mass m is at a height h above the ground and that h is small in comparison with the earth s radius so that the body s weight mg does not change appreciably with the height The gravitational potential energy E_p of the system body earth will then be mgh for this is the work done against the weight mg in lifting the body to the height h and in the absence of air resistance is the work the body can do in returning to the ground For example if a 1 kg object is 1000 meters above sea level E_p with respect to sea level is $mgh = 1\,0 \text{ kg} \times 9\,8 \text{ (m/sec}^2) \times 1000 \text{ m} = 9800$ joules But relative to a land surface of altitude say 500 meters above sea level E_p is only half as much or 4900 joules this being the work the body could do in dropping to this reference level rather than to the sea (Fig 2)

If a body is at a great distance from the earth as when a missile is fired to a very high altitude account must be taken of the change in the body s weight with distance r from the center of the earth as given by the Newtonian law of gravitation $f_g = Gmm'/r^2$ where f_g is the gravitational force of attraction G is the gravitational constant and m and m' are the masses of the body and earth respectively If the distance between the body and the earth s center is increased from r_1 to r_2 (Fig 3) the work W done against gravitational attraction and therefore the increase ΔE_p in the gravitational

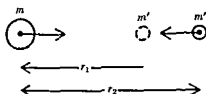


Fig 3 Finding E_p for two particles of masses m and m

potential energy of the system body earth is

$$W = \Delta E_p = \int_{r_1}^{2} f_g\, dr = E_{p2} - E_{p1} = Gmm\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \quad (2)$$

Eq (2) holds for any two bodies that can be treated as particles and therefore is applicable to the earth as one of the attracting bodies to the extent that the earth can be regarded as spherical and made up of concentric shells each of which is homogenous as regards density In the theory of gravitation the results are often simplified when the reference distance r_1 between the two attracting bodies is infinitely great that is the gravitational potential energy of a body is assumed to be zero when it is far removed from all other bodies Then from Eq (2) $E_p = -Gmm/r_2$ and since r_2 may be any distance the subscripts can be dropped giving $E_p = -Gmm/r$ The negative sign means simply that E_p at any finite distance of separation r is smaller than it is at infinity See GRAVITATION

Electrical potential energy The electrical potential energy of two particles in vacuum having charges q and q' can similarly be shown with the help of Coulomb s electrostatic law to be $E_p = \pm k_0 qq'/r$ where k_0 is a constant the algebraic sign is negative or positive according as the charges have unlike or like signs See COULOMB S LAW ENERGY SOURCES [D E R]

Bibliography R A Becker *Introduction to Theoretical Mechanics* 1954 G J Holton and D H D Roller *Foundations of Modern Physical Science* 1958

Energy level (quantum mechanics)

The energy of a stationary state The term refers to the practice of displaying energies of the stationary states in an energy level diagram on which the ground state (lowest energy state) is placed at the bottom of the diagram and on which the other (excited) stationary states are located at levels above the ground state proportional to their energy differences from the ground state In most cases but not always the energy levels are restricted to a discrete set with finite spacing that is to the bound state energies and to specially selected unbound energies such as the energies of compound states

Observed spectral lines corresponding to transitions between identified pairs of levels are also often displayed on the energy level diagram by means of (approximately) vertical lines between the appropriate levels such diagrams sometimes are called Grotrian diagrams When two or more stationary states have the same energy the corresponding level is termed degenerate The width of a level is its uncertainty in energy related to its lifetime by the requirements of the uncertainty principle *See* UNCERTAINTY PRINCIPLE *See also* ATOMIC STRUCTURE AND SPECTRA DEGENERACY (QUANTUM STATES) EXCITED STATE GROUND STATE METASTABLE STATE QUANTUM MECHANICS QUANTUM THEORY NONRELATIVISTIC, STATIONARY STATE [E G]

Energy sources

The sources from which energy can be obtained to provide heat light and power Industrial society has been based largely on the substitution for animal energy of power from heat of combustion of carboniferous fuels It seems likely that it will be based in the future largely on heat from the sun and heat which is generated by nuclear reactions Major carboniferous fuels are coal petroleum and natural gas—all fossil materials formed in finite amounts many millions of years ago Other important fossil fuels are oil shales and tar sands Minor fuels are forms of current production of vegetation Nonfuel sources of energy are water wind terrestrial heat atmospheric electricity and sunlight These last supply relatively unimportant parts of the world s total used energy but all are renewable sources of energy The supplies of elements suitable for nuclear reactions are finite but abundant

Petroleum Since 1948 imports of petroleum into the United States have been increasing at the average rate of 70 000 000 bbl a year because it has not been prudent for United States production to be made equal to United States demand During the same time demand in the rest of the world has increased more rapidly than it seems possible to increase production By the late 1950s annual world consumption of petroleum stood at over 6 000 000 000 bbl The most responsible geologic estimate for total United States reserves of petroleum both on and off shore was 81 000 000 000 bbl in 1957 and for the rest of the world about 1 275 000 000 000 bbl Only about one tenth of the total has been proved by drilling If demand trends since 1940 should continue and if geologic estimates are correct the peak of United States production will come around 1965 and the peak of production in the rest of the world around 1980 Demand will presumably still be rising when production begins its inevitable decline

Natural gas The most responsible geologic estimate of the total natural gas under United States soil is 600 000 000 000 000 cu ft of which about 40% has been proved by drilling (1957) The average increase in consumption of gas in the United States between 1945 and 1955 was 8% a year If the future rate of increase in consumption should be only 4% a year and if geologic estimates are correct the peak of production of natural gas will come between 1965 and 1970

Oil shale It is believed that the United States has about 55% of the oil shale of the world and Brazil about 43% If all United States oil shale could be converted to liquid fuel 1 000 000 000 000 bbl of oil might be produced but for reasonable cost the total might be closer to 125 000 000 000 bbl The limiting factors are richness of oil shale availability of water and facilities for disposal of ash Even from rich shale production of 1 000 000 bbl of oil per day would mean more than 1 000 000 tons of ash per day The United States Bureau of Mines estimates that facilities for the production of as much as 2 000 000 bbl of shale oil per day might be developed by the year 1980 From this is indicated a peak of shale oil production of around 1 500 000 000 bbl year about the year 2000 This would represent a small fraction of our presumed need for liquid fuel at that time

Tar sands The largest known deposit of tar is in the northern part of Alberta Canada This deposit might ultimately yield up to 500 000 000 000 bbl of oil at a fantastic cost except for about 1 000 000 000 bbl accessible to surface mining The United States is believed to have about 1% as much tar sand as Canada Tar sands can be expected ultimately to extend the world s supply of liquid fuel only a few months

Coal The United States Bureau of Mines estimates that the United States has almost 2 000 000 000 000 tons of coal of various kinds Mining engineers believe that little more than one tenth of this may be economically mined The known reserve of bituminous coal of current commercial quality and cost may be only about 27 000 000 000 tons This is readily accessible coal in beds 28 in or more thick at depths of no more than 1000 ft At probable rates of increase in demand production of this coal may reach its peak around 1970 The shortage can be made up of course by mining coal of good quality at higher cost and by using coals of poorer quality but it is doubtful that bituminous coal will be available for large scale conversion to liquid fuel About 21 000 000 000 tons of subbituminous coal and lignite can be mined economically Some of this might be used for conversion to liquid fuel

Nuclear energy Although supplies of fissionable elements are finite the amounts are so large that the timetable for production of nuclear energy depends primarily upon invention and rate of technological development These factors are even more important for use of nuclear fusion reactions The amount of energy that could be ultimately obtained from nuclear fusion is almost inconceivably large if it is found possible to generate such energy without expending large amounts of energy in the preparation of the materials used Because the data are classified it is not known how much energy is required to manufacture and maintain fissionable elements but at the present time more energy is used in the manufacture of uranium and other fuel elements than is produced in power reactors Nuclear power provides a particularly convenient form of energy for certain specific applications

Vegetation Until recent years wood was the major source of the world s energy and in most areas of the world wood is still the major fuel In 1800 the burning of wood supplied 94% of United States energy By 1885 the proportion was 50% In 1957 it was 3 5% about as much as United States hydroelectric power The proportional trend has been sharply downward because of the rapid rate of increase in total energy demand per capita and the demand for food by the world s increasing population In 1957 about one fourth of the world s vegetation production was used as fuel but this constitutes wasteful unbalance Unless sufficient vegetation is used for sustaining animal life and continued fertility of the soil and the hydrosphere vegetation must be regarded as a nonrenewable source of energy Wood together with other natural forms and known derivatives of vegetation will continue to supply a part of the world s demand for energy but a decreasing part Natural growth of vegetation is one of several ways in which solar energy can be captured Sunlight is abundant and virtually perpetual By comparison water power and wind power are of relatively little consequence

Terrestrial heat The flow of heat from the interior crust of the earth is estimated to be around 250 000 000 000 000 000 hp hr per annum 5 or 6 times the 1957 world s energy requirement The major part of this energy is believed to come from radioactive substances in a skin of surface rock not more than 20 miles deep No practical plan is in sight for the utilization of more than a trifling proportion of this energy A few power plants have been based upon the escape of volcanic heat but if all the volcanic heat of the earth could be captured it would meet only about one tenth of the world s energy requirement and much of this energy supply would be in remote islands of the Pacific Nonvolcanic earth heat has been unavailable because of the very low thermal conductivity of the earth s crust

It has been estimated that there are 20 000 miles of ocean shore line where the difference in temperature between the sun heated surface water and the cool deep water is enough to operate a condensing turbine If all engineering problems could be solved it might be possible to capture about 5% of the world s energy requirement in this way but most of this would be in areas where there is little use for energy

Atmospheric electricity Even if all the world s lightning strokes (about 300 000 000 per year of which perhaps one tenth are strokes to earth) could be combined to make an even flow of electric power it would yield less than 0 01% of United States requirements alone In addition to lightning there are several other forms of atmospheric electricity but the total amounts to only about one

of electricity

sumption of energy is dependent upon preferred devices for its use because of losses in conversion of one form of energy to another and losses in operation For example motor cars utilize gasoline which is made

from petroleum with an average loss of 13% Because of thermal mechanical and auxiliary effects the gasoline is u ed in the motor car with a further loss of around 75% Motor fuel represents the second largest United States consumption of energy (The l . . . [illegible]

. . . [illegible]

drive a generator about 22 hp-hr is obtained in the form of electric power By the time this is transmitted to points of use there remain 16 4 hp hr From this in the most efficient lamp (fluorescent), about 3.5 hp-hr is obtained in the form of light Almost one third of United States electric power is used for lighting More than half of the fuels used are subjected to conversion processes and more than half are used in devices with low efficiency The most modern nuclear reactor plants have lower thermal efficiency than modern coal plants primarily because of the use of lower pressures in the nuclear plants The reason for lack of progress in reduction of energy waste is that energy so far has been cheap and abundant See CHEMICAL ENERGY COAL NATURAL GAS, NUCLEAR POWER OIL SAND OIL SHALE, PETROLEUM, SOLAR ENERGY WATER POWER WIND POWER [E A Y]

Bibliography E. Ayres, et al Energy Sources—The Wealth of the World 1952, E Ayres The fuel situation Sci American 195(4) 43–49 1956

Engine

A machine designed for the conversion of energy into useful mechanical motion The principal characteristic of an engine is its capacity to deliver appreciable mechanical power as contrasted to a mechanism such as a clock or analog computer whose significant output is motion By usage an engine is usually a machine that burns or otherwise consumes a fuel as differentiated from an electric machine that produces mechanical power without altering the composition of matter (see MOTOR ELECTRIC) Similarly a spring driven mechanism is said to be powered by a spring motor a flywheel acts as an inertia motor By this definition a hydraulic turbine is not an engine although it competes with the engine as a prime source of mechanical power (see HYDRAULIC TURBINE PRIME MOVER WATER POWER)

Applications A fuel burning engine may be stationary as a donkey engine used to lift cargo between wharf and ship or it may be mobile like the engine in an aircraft or automobile (see AIRCRAFT ENGINE AUTOMOTIVE ENGINE) Such an engine may be used for both fixed service and mobile operation although accessory modifications that adapt the engine to its particular purpose are preferable For example the fan that draws air through the radiator of a water cooled fixed engine is large and fitted in a baffle whereas the fan of a similar but mobile engine can be small and unbaffled because considerable air is driven through the radiator by ram action as the engine propels itself along

Some types of engine can be designed for economic efficiencies in fixed service but not in mobile operation Thus the steam engine is widely used in central electric generator stations but is obsolete in mobile service This is chiefly because in a large ground installation the furnace and boiler can be fitted with means for using most of the available heat (see STEAM BOILER STEAM GENERATING UNIT see also POWER PLANT) The engine proper can be a reciprocating (piston) or a rotating (turbine) type (see STEAM ENGINE STEAM TURBINE) Because shaft rotation is by far the most used form of mechanical motion the turbine is the more common form of modern steam engine For railroad service the steam engine has given place to diesel and gasoline internal combustion engines and to electric motors (see LOCOMOTIVE)

Types Traditionally engines are classed as external or internal combustion External combustion engines consume their fuel or other energy source in a separate furnace or reactor See FURNACE (STEAM GENERATING) REACTOR NUCLEAR Strictly the furnace or reactor releases chemical or nuclear energy into thermal energy and the engine proper converts the heat into mechanical [illegible]

fluid in a boiler can be considered a combustion type (see SOLAR ENGINE) To avoid loss of or contamination from nuclear fuel the reactor and boiler are separated from (and may also be shielded from) the engine (see NUCLEAR AIRCRAFT PROPULSION REACTOR SHIP PROPULSION) The working fluid takes on energy in the form of heat in the boiler and gives up energy in the engine the engine proper being a thermodynamic device The device may be a turbine for stationary power generation on a nozzle for long range vehicular propulsion (see NUCLEAR POWER, NUCLEAR ROCKET)

In an engine used for propulsion the rearward velocity with which the working fluid is ejected and thus the forward acceleration imparted to the vehicle depend on the temperature of the fluid For practical purposes temperature is limited by the engine materials that serve to contain the chemical combustion or nuclear reaction To achieve higher exhaust velocity the working fluid may be contained by nonmaterial means such as electric and magnetic fields in which case the fluid must be electrically conductive (see ELECTROMAGNETIC PROPULSION ION PROPULSION MAGNETOGAS DYNAMICS) The engine proper is then a magnetohydrodynamic device receiving electric energy from a

conversion of nuclear or solar radiation (see INTERPLANETARY PROPULSION, THERMOELECTRICITY)

A further basis of classification concerns the working fluid If the working fluid is recirculated the er operates on a closed cycle. If the wo

ing fluid is discharged after one pass through boiler and engine the engine operates on an open cycle Closed cycle operation assures the purity of the working fluid and avoids the discharge of harmful wastes The open cycle is simpler Thus the commonest types of engine use atmospheric air in open cycles both as the principal constituent of their working fluids and as oxidizer for their fuels

If open cycle operation is used the next modification is to heat the working fluid directly by burning fuel in the fluid the engine becomes its own furnace Because this internal combustion type engine uses the products of combustion as part of the working fluid the fuel must be capable of combustion under the operating conditions in the engine and must produce a noncorrosive and nonerosive working fluid Such engines are the common reciprocating gasoline and diesel units (*see* DIESEL ENGINE INTERNAL COMBUSTION ENGINE)

At low speeds the combustion process is carried out intermittently in a cylinder to drive a reciprocating piston (*see* CARNOT CYCLE DIESEL CYCLE OTTO CYCLE) At high speed however friction between piston and cylinder walls and between other moving parts dissipates an appreciable portion of the developed power Thus where high power is developed at high speed performance is improved by continuous combustion to drive a turbine wheel (*see* BRAYTON CYCLE GAS TURBINE TURBINE PROPULSION) Engine shaft rotation may be used in the same way as in a reciprocating engine (*see* TURBOPROP) However for high velocity vehicular propulsion the energy of the working fluid may be converted into thrust more directly by expulsion through a nozzle (*see* JET PROPULSION TURBOJET TURBORAMJET) Once the vehicle is in motion the turbine can be omitted (*see* RAMJET) Alternatively instead of drawing atmospheric oxygen into the combustion chamber the engine may draw both oxidizer and fuel from storage tanks within the vehicle or the combustion chamber may contain the full supply of fuel and oxidizer (*see* ROCKET ENGINE)

Despite all the variation in structure mode of operation and working fluid—whether of moving parts moving fields or only moving working fluid—these machines are basically means for converting heat energy to mechanical energy (*see* THERMODYNAMIC PROCESSES) [F H R]

Engine cooling

Cylinder gas temperatures in internal combustion engines may reach 4500°F This is well above the melting point of the engine parts in contact with the gases so that it is necessary to control the temperature of the parts or they will become too weak to carry the gas pressure stresses The lubricating oil film on the cylinder wall can fail due to chemical changes at wall temperatures above about 400°F Complete loss of power may take place if some spot in the combustion space becomes sufficiently heated to ignite the charge early on the compression stroke

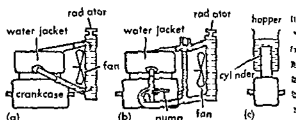


Fig 1 Engine liquid cooling systems (a) Natural circulation (b) Forced circulation (c) Hopper cooling

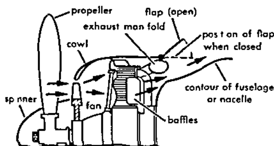
Fortunately a thin protective boundary of relatively stagnant gas of poor conductivity exists on the inner surfaces of the combustion space If the outer cylinder surface is placed in contact with a cool fluid such as air or water and there is sufficient contact area to cause a rapid heat flow the resulting temperature drop produced by the heat flow in the inside boundary layer keeps the cylinder wall temperature much closer to the coolant temperature than to the combustion gas temperature The quantity of heat that crosses the stagnant boundary layer and must be carried away by the coolant is a function of the Reynolds number of the gas existing in the cylinder In terms of practical engine quantities the heat flow to the coolant varies approximately as (charge density × piston speed)$^{0.8}$ At full throttle and normal piston speed this heat flow amounts to about 15% of the energy of the incoming fuel

Liquid cooling If the coolant is water it is usually circulated by a pump through jackets surrounding the cylinders and cylinder heads The water is circulated fast enough to remove steam bubbles that may form over local hot spots and to limit the water's temperature rise through the engine to about 15°F The warmed coolant is piped to an air cooled heat exchanger called a radiator (Fig 1) The air flow required to remove the heat from the radiator is supplied by an engine-driven fan and also by the forward motion of the vehicle in automotive applications In liquid cooling the engine and radiator may be separated and each placed in the optimum location being connected to each other through piping To prevent freezing the water coolant is often mixed with alcohol or ethylene-glycol

Low water jacket temperature is conducive to corrosive wear of the engine parts and increases the piston friction losses High water jacket temperature increases the coolant loss by evaporation or by actual boiling Temperature of the water jacket is often automatically maintained near 160°F by a thermostat placed in the line from engine to radiator When the engine outlet water is too cool it is prevented from entering the radiator and is usually recirculated in the engine block until it becomes warm enough to open the thermostat

Air cooling Engines are often cooled directly by a stream of air without the interposition of a liquid medium The heat transfer coefficient between the cylinder and air stream is much less than with a

liquid coolant so that the cylinder temperatures must be much greater than the air temperature to transfer to the cooling air the heat flowing from the cylinder gases To remedy this situation and to reduce the cylinder wall temperature the outside area of the cylinder which is in contact with the cooling air is increased by finning The heat flows easily from the cylinder metal into the base of the fins and the great area of finned surface permits heat to be transferred to the cooling air (Fig 2) The ideal fin shape depends upon the conductivity of the fin material In general the fin is thickest at the base to permit heat flow from the cylinder The fin should taper to a thin edge to give a good temperature gradient along its length For reasons of mechanical strength fins are usually made thicker than necessary for heat transfer considerations



F g 2 Air flow in a rad al aircraft eng ne

High output cylinders require many closely spaced fins In these engines the area between adjacent cylinders is blocked off with sheet metal baffles which are also shaped to follow the fin tips part way around the cylinder A pressure is built up in front of the baffles by means of a fan or because of the forward motion of the vehicle (ram effect) The pressure differential between front and rear of the engine forces the cooling air through the spaces between the fins

The power required to cool depends upon the quantity of cooling air used and the velocity at which it passes the fins For minimum cooling power the fins should be long and close together so that a large heat transfer area is served by a small coolant flow area The temperature difference between fins and cooling air should be kept as high as possible so that less air velocity will be required Cylinder temperatures of air cooled engines are sometimes controlled by louvres or flaps which may be set to restrict the cooling air flow until the engine becomes warm [A R R]

Engineering

Engineering has been defined as the art of directing the great sources of power in nature for the use and convenience of man In its modern form the practice of engineering involves men money materials machines and energy It is differentiated from science because it is primarily concerned with how to apply and direct to useful and economical ends the basic natural phenomena which scientists discover and formulate into acceptable theories Engineering therefore requires above all the creative imagination to innovate useful applications of natural phenomena It is always dissatisfied with present methods and equipment It seeks newer cheaper better means of using natural sources of energy and materials to improve man s standard of living and to diminish laborious toil

Traditional engineering Traditionally there were two divisions or disciplines military engineering and civil engineering As man s knowledge of natural phenomena grew and the potential civil applications became more complex the civil engineering discipline tended to become more and more specialized The practicing engineer began to restrict his operations into narrower channels For instance civil engineering came to be concerned primarily with static structures such as dams bridges and buildings whereas mechanical engineering split off to concentrate on dynamic structures such as machinery and engines Similarly mining engineering became concerned with the discovery of and removal from geological structures of metalliferous ore bodies whereas metallurgical engineering involved extraction and refinement of the metals from the ores From the practical applications of electricity and chemistry electrical and chemical engineering arose

This fractionating and splintering process continued as narrower specialization became more prevalent Civil engineers had more specialized training as structural engineers dam engineers water power engineers bridge engineers mechanical engineers as machine design engineers industrial engineers motive power engineers electrical engineers as power and communication engineers (and the latter eventually into telegraph telephone radio television and radar engineers whereas the power engineers divided into fossil fuel and nuclear engineering) mining engineers into metallic ore mining and fossil fuel mining (the latter into coal and petroleum engineering)

Integrating influences While this specialization was taking place there were also integrating influences in the engineering field The growing complexity of modern technology called for many specialists to cooperate in the design of industrial processes and even in the design of individual machines This brought interdisciplinary activity to coordinate the specialists For instance the design of a modern structure involves not only the static structural members but a vast complex including moving parts (elevators for example) electrical machinery and power distribution communication systems heating ventilating and air conditioning and fire protection Even the structural members must be designed not only for static loading but for dynamic loadings such as for wind pressures and earthquakes Because men and money are as much involved in engineering as materials machines and energy sources the management engineer arose as another integrating factor in modern technology

The typical modern engineer goes through several phases of activity during his engineering career His formal basic education must be broad and deep in the sciences and humanities which underlie his field Then comes an increasing degree of specialization in the intricacies of his discipline also involving continued postscholastic education Normal promotion thus brings interdisciplinary activity as he supervises the specialists under his charge Finally he enters into the management function as he interweaves men money materials machines and energy sources into completed processes for the use and convenience of man

For specific articles on various engineering disciplines *see* CHEMICAL ENGINEERING CIVIL ENGINEERING ELECTRICAL ENGINEERING MARINE ENGINEERING MECHANICAL ENGINEERING MINING NUCLEAR ENGINEERING *see also* SCIENCE TECHNOLOGY [J W B]

Engineering and architectural contracts

This article discusses some of the provisions in engineering and architectural contracts that have often been the subject matter of litigation and the interpretation of such provisions by courts of the United States

Licensing All states require that persons engaging in the profession of architecture or engineering must obtain a license A person without a license cannot recover for rendering such services This rule does not apply to persons who are employed by architects or engineers but only to those who render services to the general public

In *Palmer v Brown* 273 P 2d 306 127 Cal App 2d 44 an architectural partnership was denied recovery for work done in the name of the firm by an unlicensed member The court held that the partnership could validly contract for architectural services to be rendered by its licensed architects but could not recover for services of an unlicensed partner

The rule was also applied in a case where a construction company contracted to prepare plans and specifications for remodeling a building so that it could be used as a bar and restaurant The plans were to cover the complete job including interior decorations fittings furnishings kitchen equipment and other items The contractor was to receive $10 500 for his services The owner paid $500 and when he refused to pay more the contract r s ed He was denied recovery on the ground

[illegible]

architecture and the contract therefore was illegal *American Store Equipment & Construction Corp v Jack Dempsey s Lunch Bowl* 21 N Y S 2d 117 affirmed 283 N Y 601 (1939)

Stipulated cost contracts An architect or engineer employed under a contract that stipulates a maximum construction cost or an approximate maximum cannot exceed the stipulated or approximated limit by more than a "reasonable" margin Exceeding this margin precludes the architect from recovering for his services *Bueche v Eickenroht* 220 S W 2d 1911 (Texas Civ App 1949) This is the general rule throughout the United States In the *Bueche* case the court stated that it is a jury issue as to whether the architect has reasonably approximated the contract price If he has he may recover his total fee

This problem does not arise where the sum named is only an estimate and where the owner has required that plans and specifications respect his wishes regarding method of construction details of construction and size of the building In such a case merely exceeding the estimate does not prevent recovery of the full fee *Schuender v Schrafft* 246 Mass 543 141 N E 511 512 (1923)

Knowledge of zoning laws An architect is charged with knowledge of the zoning restrictions of a designated locality Plans he prepares for the erection of a building must comply with local ordinances If they do not he cannot recover for services rendered *Bott v Moser* 175 Va 11 7 S E 2d 217 (1940)

Ownership of plans Where an architect has prepared plans and specifications for a client and has been paid for them such plans belong to the client and not to the architect *Wright v Eisle* 83 N Y Supp 887 (1903)

Some of the standard forms of contract between architect and owner contain a provision to the effect that all drawings and specifications are the property of the architect whether the work for which they are made be executed or not

As long as the physical plans remain in the architect s possession they constitute personal property When the idea itself is not protected by patent or copyright however it has been said that there is no intrinsic property in the architect s design nor any exclusive right in the design or the reproduction When the architect has prepared plans superintended the construction of a building and has been paid for his services the plans belong to the owner

Payment for plans In one case an architect was employed to prepare plans and specifications for remodeling and repairing certain school buildings The school board of the community had fixed the architect s compensation at 8% on all work approved and let The board also reserved the right to discontinue any and all work at any time the school s interests required In that event the compensation of the architect would be determined by the schedule established by the American Institute of Architects Although it received bids the board of education did not let any contracts for the proposed work and refused for that reason to pay the architect for preparing the plans and specifications The architect sued and recovered his services *Llewellyn v Board of Education* 154 N E 889 324 Ill 254 (1926) The court held that the schedule for services established by the American Institute of Architects [illegible]

preliminary [illegible]

basic rate [illegible]

cost Upon completion of specifications and general

working drawings (exclusive of details) a sum suf ficient to increase payments on the fee to 60% of the rate or rates of commission agreed upon com puted upon reasonable cost estimated on such com pleted specifications and drawings or if bids have been received then computed upon the lowest bona fide bid or bids

In another case *Doup v Almand* 212 Ark 687 207 S W 2d 601 (1948) an architect prepared plans and specifications for a building strictly as the owner instructed The owner then came to the architect s office and for the first time indicated that he did not intend to spend more than $10 000 for the erection and that the proposed construction would exceed that amount The blueprints were ly ing on the conference table and were available to the owner but he did not request to see the plans or ask the architect to make any changes in them When the architect broached the question of com pensation for preparing the plans the owner walked out of the office The architect brought an action to recover for his services The owner said that there had been no tender of the plans and spec ifications on the part of the architect

As to the right of the architect to recover com pensation the court stated that where the architect prepares plans and specifications that substantially comply with the owner s instructions he is entitled to compensation even though for one reason or an other the building is never constructed In order to be entitled to compensation the architect must de liver the plans but if the owner refuses to accept them no tender is necessary since he is not re quired to do an idle act Under the circumstances of the case the court said that a formal tender would have been futile

In *Jones v Brisbin* 41 Wash 167 247 P 2d 891 (1952) the Supreme Court of Washington held that when prospective builders accepted the advice of architects and said nothing when changes in plans were suggested a contract for complete plans resulted even though the original negotiations indi cated an employment of a limited scope The court said in 247 P 2d at 894 Where a person with rea sonable opportunity to reject offered services takes the benefit of them under circumstances which would indicate to a reasonable man that they were offered with the expectation of compensation a contract complete with mutual assent results 1 Restatement Contracts Section 72 (1) (a)

The court held that since the contract did not fix the amount of compensation that the architects were to receive they were entitled to recover the fair and reasonable value of the services rendered even though the building was never built because the expected sources of building loans failed to ma terialize

Percentage of cost contracts An engineer con tracted with a municipality to prepare plans and specifications and render other engineering serv ices in connection with the erection of a bridge The contract provided that the engineer s compen sation was to be 7½% of the cost of the under taking

The contract for the erection of the bridge was let and as is customary the contractor posted a surety bond so that funds for completion would be available if he defaulted After the contractor had performed about one-third of the work he aban doned the job and his surety employed another con tractor to complete the work The sum paid to the original contractor plus that paid to the other con tractor to complete the work was greatly in excess of the original contract price No part of such ex cess cost was paid to the engineer for the services rendered by him

The engineer brought an action against the mu nicipality to recover 7½% of the actual cost of the work The court limited his recovery to 7½% of what it cost the city to do the work The court indi cated that the engineer s contract was with the mu nicipality and the term cost of the undertaking meant the cost to the city and did not include any excess cost paid by the surety company *Pette v City of Nashua* 50 F 2d 50 (1st Cir 1931)

Satisfactory work provision Construction con tracts usually contain a provision to the effect that the contractor s work must satisfy the architect or engineer Such a provision does not authorize the architect or engineer arbitrarily to reject material that meets specifications

In *Midgley v Cambell Building Co* 38 Utah 293 112 P 820 (1911) a plumbing contract provided that the work was to be done in accordance with certain specifications and to the satisfaction of the architect The specifications did not provide that the plumbing fixtures were to be of any particular make but the plumbing subcontractor informed the architect that he intended to use a certain make of fixture Later the subcontractor changed his mind and the architect condemned the work The subcon tractor sued The court held that the architect could not arbitrarily condemn the use of fixtures made by another concern if they were of quality equally as good as those the subcontractor had in tended to use

In certain cases where one agrees to do some thing to the satisfaction of the other party no recov ery can be had unless the other party is satisfied This principle applies however only when a ques tion of personal fancy or taste is involved In the case referred to the court held that the provision and installation of plumbing fixtures did not in volve personal fancy and that under the contract the plumbing contractor was required only to fur nish equipment that would be sanitary useful suit able and neat Beyond this the architect was not justified in rejecting them because they were not made by a particular manufacturer The general rule of law is that a contract provision to the effect that work must be done to the satisfaction of the owner or architect merely means that the work should be satisfactory to any reasonable person *Fielding & Shepley Inc v Dorr* 72 Cal App 2d 18 163 P 2d 908 (1945)

Defective plans In preparing plans and speci fications the architect must exercise ordi

in construction resulting from defective plans as his undertaking does not guarantee a perfect plan or a satisfactory result It is considered enough that the architect himself is not the cause of the failure There is no implied promise that miscalculation may not occur When however the architect does not exercise such care and skill he will be liable in damages resulting from the defects in his plans and he cannot recover compensation for preparing them *White v Pallay* 247 P 316 119 Ore 97

In *Hill v Polar Pantries* 219 S C 263 64 S E 2d 885 (1951) the court stated It seems to be well settled that where a person holds himself out as specially qualified to perform work of a particular character there is an implied warranty that the work which he undertakes shall be of proper workmanship and reasonable fitness for its intended use and if a party furnished specifications and plans for a contractor to follow in a construction job he thereby impliedly warrants their sufficiency for the purpose in view The court cited a case involving a building contract *Avent v Proffitt* 109 S C 48 95 S E 434 (1918) where an architect was held liable for failure to discover and condemn defective plastering in a house erected under his supervision If the architect also contracts to erect the building the owner is entitled to recover damages resulting from defective plans prepared by him and for defective construction work done under his supervision and for alleged fraud and misrepresentation in respect to the cost of erecting the proposed structure *Goldberg v Underhill* 95 Cal App 2d 700 213 P 2d 516 (1950)

In another case tried in the District of Columbia *Henry R Robb Inc v Urdahl* 78 A 2d 387 (1951) the owner of a garage employed engineers to prepare plans and specifications for a heating system of sufficient size to heat a building to 70°F The engineers also supervised installation Plans were prepared and a contract price of $4602 was agreed upon Upon the completion of the work it was found that the building was inadequately heated It was then discovered that the engineers in preparing the plans had made an error as a result of which one heating unit had to be relocated and three additional units installed The engineers drew up plans for additional work costing $1403 and the owner sued the engineers to recover that amount

The trial court found that if the original plans had been drawn correctly the cost of the installation would have been $183 less than the combined cost of the two jobs because the costs of labor and materials had increased between the dates of the original and the final work The court found that (1) the engineers did not guarantee that the system could be installed for any specified sum (2) the additional plans prepared by the engineers constituted a full although delayed performance of their contract and (3) the only damage suffered by the owner was the increased cost of $183 for which the engineers were held liable

Improper supervision of work When an architect assumes supervisory duties under his contract of employment he is bound to exercise care in the performance of such duties Negligence resulting in damage to the owner gives the employer a right to recover *Lindberg v Hodgens* 152 N Y Supp 229 (1915)

The purchaser of a partly erected building employed an architect to supervise the completion of the building The architect changed the plans so that some of the floor beams rested on studded partitions a violation of the local building code Defects developed in the building The owner brought an action against the architect and recovered for damages resulting from the defective construction *Straus v Buchman* 96 App Div 270 affirmed in 184 N Y 545 (1906)

In a California case *Palmer v Brown* 127 Cal App 44 273 P 2d 306 316 (1954) the court held An architect when supervision is a part of the duties assumed by him under his contract with the owner is required to exercise due care in the performance of his supervisory function and is liable to the owner for negligence on his part

Settlement of disputes Most construction contracts provide that disputes arising out of the work shall be referred to the architect or engineer and his decision shall be considered binding and conclusive and that when the architect issues a certificate that the work has been satisfactorily completed the contractor has a right to receive the balance due him under the contract The dispute that arises most often is over a claim by a contractor for work he has been directed to perform although it is not in the contract and for which the architect or engineer refuses to issue a certificate of payment

The general rule of law is that the final certificate of the engineer or architect is conclusive unless fraud bad faith or an obvious mistake can be shown It is equally established that the final certificate is not binding upon the contractor when the engineer has erred in his legal interpretation of the contract *Uvalde Contracting Co v City of New York* 160 App Div 284 145 N Y Supp 604 (1914)

In a New Jersey case *Terminal Construction Corp v Bergen County Hackensack River Sanitary Sewer District Authority* 18 N J 294 113 A 2d 787 (1955) where this issue was involved the court held that the engineer s acts in the exercise of the authority vested in him under the terms of the contract are legally fraudulent only if arbitrary and unreasonable This rule applies even where the owner is not a direct participant in the engineer s fraud The term fraud as applied to a construction contract has a broader meaning than usual since it includes arbitrary action or gross mistake

"It has been said" the court noted "that the architect or engineer occupies a position of trust and confidence and that he should act in absolute and entire good faith throughout and when he acts under a contract as the official interpreter of its con

ditions and judge of its performance, he should side neither with the Owner or Contractor, but exercise impartial judgment"

In *Smith Contracting Co v City of New York* 240 N Y 491 (1925), the New York Court of Appeals held that a contractor is entitled to recover payment for work done (1) when the certificate for payment has been arbitrarily withheld (2) when it has been withheld as a result of an erroneous interpretation of the law, or (3) when the decision of the engineer is patently erroneous Under these circumstances it is as though the engineer were acting in bad faith

Defects in contractor's work. In the absence of any express warranty or negligence on his part, a contractor is not liable for defects that result from defective plans and specifications This principle of law was well expressed in a Mississippi case, *Trustees of the First Baptist Church v McElroy*, 223 Miss 327, 78 So 2d 138 (1955), where a contractor was sued for damages from an explosion of a gas operated steam generator in the chimney of a building The owner blamed the explosion on faulty installation but the trial court held that the work had been done to plans and specifications and approved by the supervising architect The judgment of the trial court was affirmed by the appellate court which stated "The law has become well settled in practically every American jurisdiction in which the matter has been involved that a construction contractor who has followed plans and specification furnished by the contractee, his architect or engineer and which have proved to be defective or insufficient will not be responsible to the contractee for loss or damage which results from the defective or insufficient plans or specifications, in the absence of negligence on the contractor's part or any express warranty by him as to their being sufficient or free from defects If any dangerous condition existed in connection with the vents installed by the contractor it resulted from plans and specifications prepared by the owner's architect and which the contractor was required to follow by the terms of the contract Some of the cases ground the rule on assumption of risk by the owner or contractee The majority of the cases based it on an implied warranty by the owner that the plans or specifications are suitable for the particular purpose coupled with an absence of any express warranty by the contractor in regard to the sufficiency of the specifications"

In another case *Helm v Speith* 298 Ky 225 182 S W 2d 635 (1944) a contract for the construction of a building provided that the contractor would guarantee the cellar would be watertight Quoting from 9 *Am Jur* 10 the court stated "A provision in a contract for construction of a cellar according to specifications 'the whole to be watertight and guaranteed' binds the contractor only so far as his own work is concerned and does not guarantee the plans will produce a watertight cellar"

Right to vary quantity of work. Most construction contracts give the owner the right to vary the quantity of work to be done The courts say such provisions cover only incidental changes necessary to complete the work and that a radical change constitutes a breach of contract

This principle of law was discussed by the courts in the case of *Drainage District No 1 v Rude*, 21 F 2d 257 (8th Cir 1927) In this case the proposal for bids called for installation of approximately 16 miles of tile drain but only about 7½ miles were actually installed The court held that the contract provision authorizing an increase or decrease in the quantity of work to be done did not authorize such a radical change

The court quoted one of its earlier decisions, 21 F 2d at 261 "The customary provisions in such contracts that the corporation or its engineers may make any necessary or desirable alterations in the work, and that the contractors shall receive the contract price or a price fixed by the engineer for the work or materials required by the alteration, is limited in the same way by the intention of the parties when the contract was made to such modifications of the work described in the contract as do not radically change its nature or its cost Material quantities of work required by such alteration, that are substantially variant in character and cost from that contemplated by the parties when they made their agreement constitute new and different work not governed by the agreement for which the contractor may recover its reasonable value"

The court continued at 262 stating that it was reluctant to extend the meaning of such provisions because "all safeguards thrown around the making of such contracts become futile the contract can be let at an excessive figure because no contractor can tell how many men or how much equipment or material to furnish " [I V W]

Bibliography E R Dillavou and L P Simpson, *Law for Engineers and Architects*, 4th ed, 1958, C W Dunham and R D Young, *Contracts Specifications and Law for Engineers* 1958 I V Werbin, *Legal Cases for Contractors, Architects and Engineers*, 1955, I V Werbin *Legal Guide for Contractors, Architects and Engineers*, 1952, I V Werbin *Legal Phases of Construction Contracts*, 1946

Engineering drawing

A graphical language used by engineers and other technical personnel associated with the engineering profession The purpose of engineering drawing is to convey graphically the ideas and information necessary for the construction or analysis of machines structures or systems

The basis for most engineering drawing is orthographic representation (projection) *See* DESCRIPTIVE GEOMETRY Objects are depicted by means of front, top side auxiliary, or oblique views, or combinations of these views The complexity of an object determines the number of views shown At times pictorial views are also shown (*see* P . DRAWING)

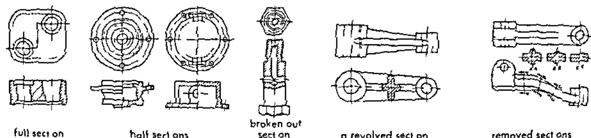


Fig. 1. Sectional views. (*From T. E. French and C. J. Vierck, Engineering Drawing, McGraw-Hill, 1953*)

Engineering drawings often include such features as various types of lines, dimensions, lettered notes, sectional views, and symbols. These drawings may be in the form of carefully planned and checked mechanical drawings (*see* DRAFTING). They may also be freehand sketches. Usually a sketch precedes the mechanical drawing. Final drawings are usually made on tracing paper or cloth so that many copies can be made quickly and cheaply by such processes as blueprinting, ammonia developed (diazo) printing, or lithography. *See* PHOTOCOPYING PROCESSES.

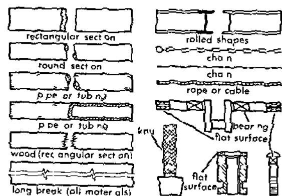


Fig. 2. Conventional breaks and other symbols. (*From T. E. French and C. J. Vierck, Engineering Drawing, McGraw-Hill, 1953*)

Many objects have complicated interior detail which cannot be clearly shown by means of front, top, side, or pictorial views. Sectional views enable the engineer or detailer to show the interior detail in such cases. Features of sectional drawings are cutting-plane symbols, showing where imaginary cutting planes are passed to produce the sections, and section lining (sometimes called cross-hatching) which appears in the sectional view on all portions that have been in contact with the cutting plane. When only a part of the object is to be shown in section, such conventional representation as a revolved, rotated, or broken-out section is used (Fig. 1). Details such as flat surfaces, knurls, and threads are treated conventionally, which facilitates the making and reading of engineering drawings by experienced personnel (Fig. 2).

In addition to describing the shape of objects, many drawings must show dimensions so that workmen can fabricate parts that will fit together. This is accomplished by placing the required values

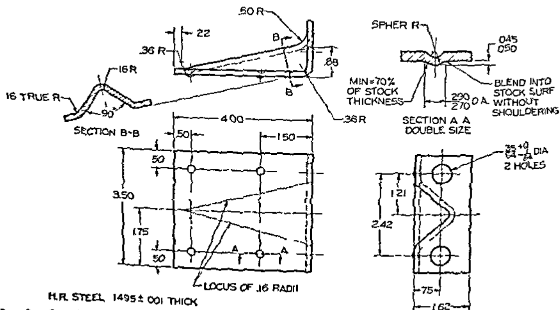


Fig. 3. Complete decimal dimensioning. (*Chevrolet*)

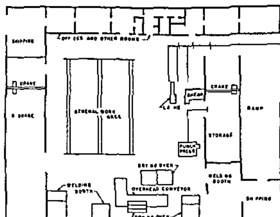


Fig 4 A plant layout drawing (From F Zozzora, *Engineering Drawing* McGraw Hill, 1958)

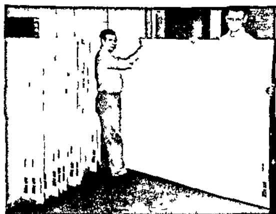


Fig 5 Aircraft master layouts photographed on steel sheets (Boeing)

(measurements) along dimension lines (usually but not always outside the outlines of the object) and by giving additional information in the form of notes which are referenced to the parts in question by angled lines called leaders (Fig 3)

Working types of drawings may differ in styles of dimensioning lettering (inclined lower case vertical
numbers
numbers
types of fractions (common fractions or decimal fractions) If special precision is required an upper and a lower allowable limit are shown Such tolerance or limit dimensioning is necessary for the manufacture of interchangeable mating parts but unnecessarily close tolerances are very expensive See DIMENSIONING

Layout drawings of different types are used in different manufacturing fields for various purposes One is the plant layout drawing in which the outline of the building work areas aisles and individual items of equipment are all drawn to scale (Fig 4) Another type is the aircraft or master layout which is drawn on glass cloth or on steel or *aluminum sheets* The object is drawn to full size with extreme accuracy The completed drawing is photographed with great precision, and a glass negative made From this negative, photo templates are made on photosensitized metal in various sizes and for different purposes thereby eliminating the need for many conventional detail drawings (Fig 5) Another type of layout, or *preliminary* assembly drawing is the design layout which establishes the position and clearance of parts of an assembly

A set of working drawings usually includes detail drawings of all parts and an assembly drawing of the complete unit Assembly drawings vary somewhat in character according to their use, as (1) design assemblies or layouts (2) working drawing assemblies (3) general assemblies (4) installation assemblies and (5) check assemblies A typical general assembly may include judicious use of sectioning and identification of each part with a numbered balloon (Fig 6) Accompanying such a drawing is a parts list in which each part is listed by number and briefly described the number of pieces required is stated and other pertinent information given Parts lists are best placed on *separate sheets and typewritten to avoid time consuming* and costly hand lettering

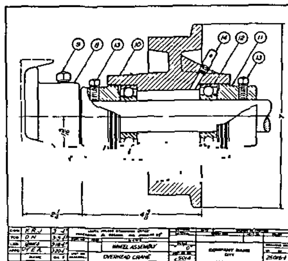


Fig 6 A unit or general, assembly (From T E French and C J Vierck *Engineering Drawing*, Mc Graw Hill 1953)

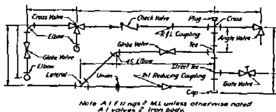
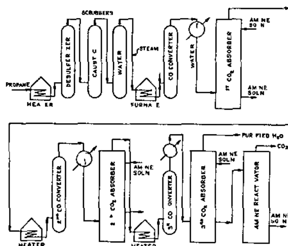
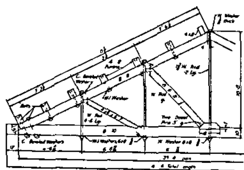


Fig 7 Piping diagrammatic drawing (From T E French and C J Vierck, *Engineering Drawing* Mc Graw Hill 1953)



Fg 8 Chemical engineering flow diagram (*From P H Groggins ed Unit Processes in Organic Synthesis 5th ed McGraw Hill 1958*)



Fg 9 Structural drawing of timber truss (*From T E French and C J Vierck Engineering Drawing McGraw H ll 1953*)

Sch-- - ..
of sta
bols
ing and electrical schematic diagrams symbols recommended by the American Standards Association (ASA) are used (Fig 7)

Additional information is often lettered on schematic drawings as for example the identification of each replaceable electrical component (*see* SCHEMATIC DRAWING) Etched circuit drawing has revolutionized the wiring of electronic components By means of such drawing the wiring of an electronic circuit is photographed on a copper clad board and unwanted areas are etched away On electrical and other types of flow diagrams all single lines (often with arrows showing direction of flow) are drawn horizontally or vertically with but few exceptions (Fig 8) Each symbol of a flow diagram must be identified preferably with the lettering placed within the enclosure

dr
da
f

ings embody the same principles as do other engineering drawings but use terminology and dimensioning techniques different from those shown on previous illustrations This system also includes special symbols which if understood facilitate the reading of a structural drawing Terminology is somewhat different in that top and front views for example become plan and front elevation in this type of drawing *See* ENGINEERING GRAPHICS WIRING DIAGRAM *See also* GRAPH THEORY, GRAPHIC METHODS, NOMOGRAPH, TOPOGRAPHIC SURVEYING AND MAPPING [A S P C J B]

Bibliography T E French and C L Svensen *Mechanical drawing* 6th ed 1957 T E French and C J Vierck *A manual of engineering drawing* 8th ed 1953 T E French and C J Vierck *Graphic Science* 1958 F E Giesecke A Mitchell and H C Spencer *Technical drawing* 4th ed 1958 R P Hoelsher and C H Springer, *Engineering drawing and geometry* 1956 F Zozzora *Engineering drawing* 2d ed 1958

Engineering geology

The solution of geological problems that arise in planning construction and maintenance of civil engineering structures by using proper data from geology and other earth sciences The branches of geology and other earth sciences most applicable in engineering geology are (1) surficial geology (2) petrology or study of rock properties (3) geohydrology or study of subsurface water and (4) geophysics particularly seismology or the study of earthquakes and their effects on engineering structures This article discusses some of the practical aspects of engineering geology

Geotechnics is the integration of pertinent geological and other earth science data with elements of civil engineering to provide a framework of information that will aid the engineer and geologist in the solution of problems connected with the natural environment of an engineering structure particularly the surrounding ground

Engineering properties of rock To the engineer and the engineering geologist most hard and compact natural materials of the earth crust are referred to as rocks whereas their derivatives formed mostly by weathering processes are soils Although a number of useful soil classification systems exist there are no convenient rock classification systems for engineering purposes There fore geological rock classification systems based primarily on the origin of the rock rather than on its significant engineering properties have to be used in engineering geology *See* ROCK SOIL MECHANICS

Rock sampling The properties of a rock can be determined by tests on samples extracted from boreholes Usually these holes are sunk by one or a combination of the following basic types of drills (1) the rotary or core drill (2) the cable tool or churn drill and (3) the auger The rotary type is used for rock In the rotary rig a motor drives a drill head which rotates usu

ally a thick walled hollow pipe) with a bit at the end The rock is cut chipped and ground The extraction of cylindrical rock cores from the bottom of a bore hole is accomplished by attaching a single or double tube core barrel to the bottom of the drill rod A bit containing fixed or replaceable reaming diamonds or other cutting teeth is screwed onto the end of the barrel *See* BORING AND DRILLING MINERAL.

Compressive strength The compressive (crushing) strength of rocks is measured in pounds per square foot (psf) or per square inch (psi) It is the stress required to break a loaded sample unconfined at the sides (Fig 1) If the load $P = 40\,000$ lb is applied to a sample with a cross section of 2 in by 2 in (or 4 sq in) the compressive stress is $40\,000 - 4 = 10\,000$ psi If this load breaks the sample the ultimate compressive strength of the rock equals the compressive stress acting at the moment of failure in this case 10 000 psi Unweathered igneous rocks particularly basalts some quartzites and siliceous cemented sandstones have the highest compressive strengths and are excellent structure foundations In a sedimentary rock the compressive strength generally depends on the quality of its cement (clay cement gives low strength) and the amount of water in the rock (increasing water content generally decreases the strength) The accompanying table shows the results of several compression tests

The presence of fissures and seams in the rock is detrimental to the compressive strength especially if the direction of these fissures coincides with the failure planes The compressive strength also depends on the direction of the acting compressive stress with relation to bedding that is the highest compressive strength is obtained when the compressive stress is normal to the bedding Conversely the highest Young s modulus of elasticity (E) generally is obtained when the compressive stress parallels the bedding

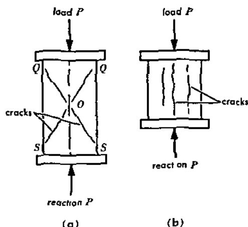


Fig 1 Unconfined compression test (a) Shear failure showing failure planes QS (b) Tension failure (From D P Krynine and W R Judd Principles of Engineering Geology and Geotechnics McGraw Hill 1957)

Compressive strength of rocks[a]

Type of rock	No of tests[b]	Compressive strength, psi (averages)
Andesite[c]	3 3	18 710–19 150
Basalt	3 3	24 450–31 850
Gneiss[c]	1 1	9310–15 140
Granite[d]	12	33 200
Granite slightly altered[c]	3 3	8250–9400
Limestone[d]	12	10 900
Limestone reef breccia	2 4	860–4960
Marble[d]	12	30,800
Sandstone (two types)[d]	12 12	10 400 and 6100
Sandstone[c]	9 8	8810–12 200
Schist, biotite	1 1	7750–12 010
Schist, biotite-chlorite	1 1	5290–17 000
Schist biotite-sillimanite[c]	1 1	1160–4930
Schist biotite sillimanite-quartz[c]	1 2	1250–4520
Slate[d]	12	30 400
Tuff	3	530

[a] Adapted from D P Krynine and W R Judd *Principles of Engineering Geology and Geotechnics* McGraw Hill, 1957

[b] First figure refers to number of specimens tested to arrive at first average value under compressive strength second figure refers to number of specimens tested to arrive at second average value

[c] Rocks tested by U S Bureau of Reclamation

[d] Rocks tested by U S Bureau of Mines

Shear in rocks Shearing stresses tend to separate portions of the rock (or soil) mass Faults and folds are examples of shear failures in nature In engineering structures every compression is accompanied by shear stresses An arch dam compresses the abutment rock if the latter is intersected by fissures it may fail in shear causing pulling and fissuring of the concrete at the dam s center The application of loads over long periods of time on certain rocks may cause them to creep and even to flow as a dense fluid (plastic flow) *See* ROCK MECHANICS STRUCTURAL GEOLOGY

Residual stress This type of stress in confined rock is actually potential energy created by ancient natural forces recent seismic activity or nearby man made disturbances Residual or stored stress may remain in rock a long time after the disturbance is removed An excavation such as a tunnel or quarry will relieve the rock by permitting displacements and thus causing conversion of the potential energy to kinetic energy In tunnels this often results in violent movement of the rock walls and floor

The transmittal of the overburden weight to the tunnel sides particularly in unlined tunnels may be explained by the phenomenon known as arching (Fig 2) Thus nature by developing self balanced systems of shearing stresses establishes a new state of equilibrium in the material around the opening An appraisal of the arching capacity of the rocks around a proposed tunnel is an important aspect of the preliminary study preceding construction Massive igneous rocks generally offer favorable arching possibilities In lined tunnels the lining carries a portion of the overburden weight *See* TUNNEL

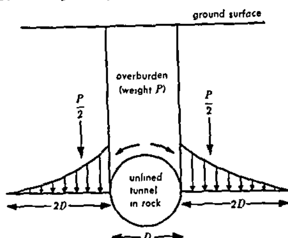


Fig 2 Arching around an opening in rock Weight P of the overburden is transmitted to the sides of the opening One half the weight of the overburden $P/2$ is distributed along a variable distance on both sides of the opening In practice this distance is generally approximated by $2D$ where D is the diameter of the opening

Construction material As a construction material rock is used in the form of dimension crushed or broken stone (*see* STONE AND STONE PRODUCTS) Broken stone is placed as riprap on earth slopes to protect them against water action Dimension stone (granite limestone sandstone) consists of blocks of predetermined shape and size that are used mostly for facing expensive buildings Crushed stone (primarily limestone but also some basalt granite sandstone and quartzite) is used as concrete aggregate Alkalies (sodium and potassium oxides) released in the chemical reactions set up by the cementing action as the concrete sets attack concrete aggregate made of deleterious material such as (1) opal and chalcedony (2) volcanic rocks containing glass devitrified glass and tridymite (3) phyllites containing hydromica (illites) and (4) other rocks containing free silica (SiO_2) Preliminary petrographic analyses of the aggregate and chemical analysis of the cement can indicate the possibility of alkali reactions and thus prevent construction difficulties such as expansion cracking or strength decrease of the concrete *See* CONCRETE PETROGRAPHY

Geotechnical significance of soils. Glacial and alluvial deposits contain heterogeneous mixtures of pervious (sand gravel) and impervious (clay silt, rock flour) soil materials The pervious materials are widely used for highway subgrade and concrete aggregate Dam reservoirs containing such deposits may be endangered by the presence of hidden pervious beds Deep alluvial deposits in or close to river deltas may contain very soft materials such as organic silt or mud *See* DELTA, FLOOD PLAINS

Eolian (wind blown) deposits Loess a porous mixture primarily of silt and fine sand has a vertical permeability considerably greater than the horizontal When a loaded loess deposit is wetted it rapidly consolidates and the overlying structure settles Permanently dry loess however, is a relatively strong bearing material *See* LOESS

Sand deposits In regions where sand dunes are present the major problem is the stabilization of the movable sand This is done by planting certain grass varieties such as heather, young pine, and other suitable plants or by treating with crude oil Cuts (excavations) are traps for moving sand and should be avoided

Organic deposits In terrains characterized by swamps peat deposits, and muskeg the major problem is building stable road embankments Good drainage avoiding cuts removal of the organic material and its replacement by sand and gravel are recommended Heavy structures are built on piles

Residual soils These soils are derived from the in place deterioration of the underlying bedrock Of least engineering concern is the generally thin granular layer derived from granite, of major concern is the irregular compressible clay deposit derived from underlying limestone

Clays supporting structures may slowly consolidate over a long period of time and thus cause settlement of structures or may expand if wetted (montmorillonites), silts may settle rapidly under a load or offer a "quick" (soft or semifluid) condition if saturated, and loosely deposited sands, often very pervious may rapidly consolidate when loaded especially in the presence of water or vibration *See* SOIL MECHANICS

Geotechnical investigation. Geotechnical investigations utilized in engineering projects may include preliminary studies preconstruction investigations and consultation and supervision during construction

Preliminary studies Preliminary studies commonly are made to select the best location for a project and to aid in planning the construction operation (*see* CONSTRUCTION ENGINEERING)

The preliminary study may entail a library search for geologic literature on the general region Maps and explanatory texts can be found in publications of governmental agencies (for example U S Army Corps of Engineers U S Geological Survey), of state agencies, of private companies, such as oil companies, and in student theses at local universities

Topographic maps and air photos can be used to study rock outcrop and drainage patterns land forms geologic structure nature of soil and vegetative cover and land use by man *See* AERIAL PHOTOGRAPH, TOPOGRAPHIC SURVEYING AND MAPPING

Field reconnaissance may include the collection of rock and soil samples the inspection of road cuts and other excavations and the inspection of nearby engineering structures Sources of construction material should be noted Aerial inspection often is helpful

Preconstruction Surface and subsurface investigations are important prior to construction Surface studies involve the preparation of a detailed

surficial geologic map showing topography hydrologic features (streams springs swamps) jointed and faulted areas and well defined landforms

Subsurface investigations are used to confirm and to amplify the preliminary geologic data These may include test pits or short tunnels (drifts) and the drilling of vertical horizontal and oblique boreholes Their locations should be plotted on a grid map of the area subdivided into numbered squares The borehole camera is a useful device for photographing the walls A later modification is a television viewing device

Data obtained during the investigations are recorded on log forms and may include rock type drilling difficulties core recovery permeability of the rock and depth of water table (with date of observation) Geophysical exploratory methods can be used to advantage if the results are checked against the data obtained by actual borings

Construction Geotechnical supervision is required *during the construction of deep excavations* and in most tunnel work The engineering geologist must advise and keep a record of all geotechnical difficulties encountered during the construction stage After construction geotechnical problems pertinent to operation and maintenance of the project often require the services of the engineering geologist

Special geotechnical problems In arctic zones structures built on permafrost may be heaved or may cause thawing and subsequent disastrous settlement In the planning of harbors and reservoirs sedimentation studies are made because soil carried by moving water will settle out and block or fill such features In seismic areas aseismic (earthquake proof) design requires knowledge of earthquake forces to construct buildings and dams that will safely resist severe earth shocks Prevention and rehabilitation of slides (landslides) in steep natural slopes and in excavations are important considerations in many construction projects *See* EARTHQUAKE FROST ACTION LANDSLIDE PERMAFROST SEDIMENTATION (GEOLOGY)

[D P K W R J]

Bibliography E C Dapples *Basic Geology for Science and Engineering* 1959 Geol Soc Am Eng Geol Div *Engineering Geological Case Histories* no 1 1957 no 2 1958 no 3 1959 D P Krynine and W R Judd *Principles of Engineering Geology and Geotechnics* 1957 J L Savage *Application of Geology to Engineering Practice* Berkey vol 1950

Engineering graphics

Any method that includes drawing lines on sheets usually paper or cloth used in the design of the products of engineering and in the computation analysis and presentation of engineering data

Engineering graphics includes preparation of drawings that show the shape of objects (*see* DESCRIPTIVE GEOMETRY PICTORIAL DRAWING) The shape may be depicted for purposes of description and explanation or for a more specific purpose such as to guide production personnel in a wire room (*see* WIRING DIAGRAM) All such diagrams are laid out in accordance with established practices the more permanent ones are drawn carefully with instruments others may be drawn freehand (*see* DRAFTING ENGINEERING DRAWING) Such graphic methods serve principally to record and communicate the geometrical aspects of products The drawings provide information for the fabrication of parts assembly of devices and maintenance of equipment To avoid laborious repetition of insignificant detail many drawings use simplified symbols in their representation (*see* SCHEMATIC DRAWING)

Graphical constructions such as descriptive geometry serve in addition as a means for arriving at a desired design Similarly graphical presentation of data aids in their interpretation (*see* GRAPH THEORY GRAPHIC METHODS) Many computations can be performed graphically other repetitive calculations *for which approximate numerical answers* are adequate can be performed graphically (*see* NOMOGRAPH SLIDE RULE) In those cases where the engineering analysis is primarily geometric as in the resolution of forces and vectors graphic constructions may be more direct than other methods Where motions of parts determine their shapes drafting techniques may even produce a design faster than analytical methods [A S P C J B]

English walnut

A large deciduous tree (*Juglans regia*) and its fruit a true nut It is also called Persian walnut and is grown throughout temperate and subtropical climatic zones of the world The nuts are used for food by man and the wood is valued for furniture and paneling

The English or Persian walnut is native to Asia Minor and was first brought to North America from England It is best adapted to areas with a fairly mild winter climate such as southern California where about 150 000 acres are planted The Mediterranean Basin countries and India China South Africa and Australia also produce English walnuts in volume Some species from the Carpathian



English walnut (a) Twig with leaves and fruit (b) Hulled n

Mountains of Europe withstand temperatures of 20°F or lower The nuts high in protein and fat, are consumed mainly as a dessert nut as an ingredient in various confections cakes cookies and other bakery goods in ice cream and to some extent as a meat substitute *See* NUT CROP CULTURE [E F S]

Enopla

A class or subclass of the phylum Rhynchocoela which is divided into two orders the Bdellomorpha (Bdellonemertini) and Hoplonemertini The proboscis of the hoplonemertines is armed whereas



Diagrammatic transverse section of a hoplonemertine The different groups of nemertines are differentiated by the number of rings of longitudinal muscles and the position of the nerve cords (*From L A Borradaile F A Potts et al The Invertebrata 3d ed Cambridge 1958*)

that of the bdellonemertines is unarmed The mouth opening is anterior to the brain and the nervous system lies below the musculature of the body wall *See* ANOPLA BDELLOMORPHA HOPLONEMERTINI [C B C]

Enoploidea

Free living nematodes with pocketlike amphids a simple cuticle 16 sense organs and a simple straight esophagus This group of animals is regarded as a superfamily by most nematologists however some authorities give it the status of an order It includes many of the larger free living nematodes but most are a millimeter or less in length Two families are mainly marine, one of these the *Oncholaimidae* is unusual in that it has a demanian system This is a double system of ducts of unknown function extending between the intestine the reproductive organs, and lateral pores in the body wall Discovered by de Man in 1884 it was designated the 'Demanian System' by H A Cobb in 1930 Four other families sometimes grouped under the Tripyloidea occur in moist soil and in fresh and brackish water One of these the Mononchidae and other enoploids including the Oncholaimidae are predatory the remainder feed on algae diatoms and bacteria The Dorylaimoidea and Mermithoidea were formerly grouped as families of this superfamily *See* NEMATODA [H E W]

Enstatite

The name given to the magnesian end member ($MgSiO_3$) of the orthorhombic pyroxene solid solution series (*see* PYROXENE) The mineral is characterized by the (110) cleavages 87° apart It is usually of a yellowish gray color, becoming greenish with a little iron present transparent in thin sections optically positive with refractive indices $n\alpha = 1.650$ $n\beta = 1.653$ and $n\gamma = 1.658$ The calcium free enstatite inverts at approximately 990°C to protoenstatite (also orthorhombic) Above 990°C protoenstatite is the stable form up to the decomposition temperature On rapid cooling below 990°C protoenstatite inverts to a metastable monoclinic form called clinoenstatite However clinoenstatite becomes the stable high temperature form if small amounts of calcium are in solid solution in the mineral

Enstatite is a common constituent in many basalts dunites serpentinites and peridotites and occurs in slags and meteorites Olivine diopside calcium pyroxenes and calcium rich plagioclases are common associated minerals Single crystals of enstatite altered to a single crystal of antigorite (serpentine) called bastite that preserves the pyroxene crystal outline are frequent in certain serpentine masses Enstatite alteration to amphibole is also common *See* PIGEONITE [C W D]